

بهبود عملکرد الگوریتم خوشه‌بندی فازی سی - مینز با وزن دهی اتوماتیک و محلی ویژگی‌ها

سمیرا رفیعی^۱، دانشجوی کارشناسی ارشد، پرهام مرادی^۲، استادیار

۱- گروه مهندسی کامپیوتر و فناوری اطلاعات - دانشکده مهندسی - دانشگاه کردستان - سنندج - ایران - s.rafaee@eng.uok.ac.ir

۲- گروه مهندسی کامپیوتر و فناوری اطلاعات - دانشکده مهندسی - دانشگاه کردستان - سنندج - ایران - p.moradi@uok.ac.ir

چکیده: خوشه‌بندی یکی از راهکارهای مهم در تحلیل داده‌هاست. در راهکارهای کلاسیک خوشه‌بندی معمولاً فرض بر این است که همه ویژگی‌ها از درجه اهمیت یکسانی برخوردار هستند. این در حالی است که در مجموعه داده‌های واقعی یک سری از ویژگی‌ها اهمیت بیش‌تری نسبت به دیگر ویژگی‌ها دارند. در نتیجه ویژگی‌های مهم تأثیر بیش‌تری در شناسایی خوشه‌های بهینه نسبت به سایر ویژگی‌ها خواهند داشت. در این مقاله، یک الگوریتم خوشه‌بندی فازی با وزن دهی اتوماتیک محلی ارائه شده است. الگوریتم پیشنهادی دارای سه مزیت اصلی است. مزیت اول این است که وزن دهی ویژگی‌ها را به صورت محلی انجام می‌دهد به طوری که وزن ویژگی‌ها در یک خوشه نسبت به خوشه دیگر متفاوت است. مزیت دوم، استفاده از یک معیار شباهت غیراقلیدسی برای به دست آوردن فاصله بین نمونه‌ها به منظور کاهش حساسیت الگوریتم نسبت به نویز است. مزیت سوم این است که وزن ویژگی‌ها به صورت تطبیقی و در طول فرآیند یادگیری به دست می‌آید. در این مقاله تحلیل‌های ریاضی برای به دست آوردن توابع به‌روزرسانی مراکز خوشه‌ها و وزن‌های ویژگی‌ها ارائه شده است. همچنین تحلیل‌های ریاضی جهت اثبات همگرایی الگوریتم نیز ارائه شده است. آزمایش‌های انجام گرفته بر روی یک مجموعه داده مصنوعی و پنج مجموعه داده واقعی نشان‌دهنده کارایی بالای الگوریتم پیشنهادی در مقایسه با دیگر الگوریتم‌های پیشنهادی با وزن دهی سراسری و محلی ویژگی‌ها است.

واژه‌های کلیدی: خوشه‌بندی فازی سی - مینز، وزن دهی محلی ویژگی‌ها، معیار شباهت غیراقلیدسی.

Improving Performance of Fuzzy C-means Clustering Algorithm using Automatic Local Feature Weighting

S. Rafiee¹, MSc Student; P. Moradi², Assistant Professor

1- Department of Computer Engineering and Information Technology, Faculty of Engineering, University of Kurdistan, Sanandaj, Iran, Email: s.rafaee@eng.uok.ac.ir

2- Department of Computer Engineering and Information Technology, Faculty of Engineering, University of Kurdistan, Sanandaj, Iran, Email: p.moradi@uok.ac.ir

Abstract: Clustering is one of the most widely used methods in data analysis. In the classical clustering methods it is assumed that all the features of the samples in a given data set make equal contribution when constructing the optimal clusters. However, in real-world datasets, some of the features can exhibit higher relevance than others. Thus, the features with higher relevance are more important to form optimal result than those with lower relevance. To address this issue, in this paper we proposed a feature weighted fuzzy clustering method called RLWFCM. The proposed algorithm has three main advantages. The first advantage is that our method assigns different weights to each feature in each cluster. This is due to the fact that each feature may have different importance in different clusters. The second advantage is using a specific kernelized distance to reduce the noise-sensitivity of the algorithm. The third one is using an analytical method for adapting feature weights during training process. Moreover, in this paper the mathematical analysis to proof the convergence of the algorithm has been presented. Several experiments were performed on an artificial dataset and also five real world datasets and the results show that the proposed method outperformed the other feature weighted clustering methods.

Keywords: Fuzzy C-Means clustering, local feature weighting, non-euclidean similarity measure.

تاریخ ارسال مقاله: ۱۳۹۳/۱۱/۲۸

تاریخ اصلاح مقاله: ۱۳۹۳/۱۲/۲۷

تاریخ پذیرش مقاله: ۱۳۹۴/۰۴/۰۹

نام نویسنده مسئول: پرهام مرادی

نشانی نویسنده مسئول: ایران - کردستان - سنندج - بلوار پاسداران - دانشگاه کردستان - دانشکده مهندسی - گروه مهندسی کامپیوتر و فناوری اطلاعات.

۱- مقدمه

از ایده استفاده شده در مقاله [۲۶]، از فاصله‌ای غیر از فاصله اقلیدسی استفاده شده است. مزیت استفاده از این معیار فاصله این است که به ویژگی‌های داده‌هایی که دارای فاصله بیش‌تری نسبت به سایر نمونه‌ها هستند (داده‌های پرت^۱) وزن کم‌تری اختصاص داده و در نتیجه این داده‌ها تأثیر کم‌تری بر فرآیند خوشه‌بندی خواهند داشت.

برای بررسی کارایی الگوریتم پیشنهادی، آزمایش‌های مختلفی بر روی پنج مجموعه داده از انبار داده UCI^۲ انجام گرفت. همچنین الگوریتم پیشنهادی با الگوریتم‌های فازی سی- مینز با وزن‌دهی اتوماتیک سراسری [۱۹] و هارد سی- مینز با وزن‌دهی اتوماتیک محلی [۲۵] مقایسه شد. نتایج به‌دست‌آمده نشان‌دهنده کارایی بالای الگوریتم پیشنهادی نسبت به الگوریتم‌های مذکور است.

ادامه مقاله به این صورت سازمان‌دهی شده است که در بخش ۲ جزئیات الگوریتم پایه فازی سی- مینز [۲۷] ارائه شده است. در ادامه در بخش ۳ الگوریتم پیشنهادی معرفی شده و بخش ۴ شامل نتایج آزمایش‌ها است و در نهایت در بخش ۵ به نتیجه‌گیری و کارهای آتی پرداخته خواهد شد.

۲- الگوریتم فازی سی- مینز

در این بخش الگوریتم پایه فازی سی- مینز [۲۷] به‌طور خلاصه معرفی خواهد شد. تابع هدف این الگوریتم به‌صورت زیر تعریف می‌شود:

$$J = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m d^2(x_i - v_j) \quad (1)$$

که μ_{ij} میزان تعلق نمونه x_i به مرکز خوشه v_j و m درجه فازی بودن را تعیین می‌کند. در اینجا $d^2(x_i - v_j)$ همان فاصله اقلیدسی و برابر با $(x_i - v_j)^2$ است. به‌طوری‌که x_i نمونه i ام و v_j مرکز خوشه j ام است. برای این تابع هدف، محدودیت‌های $0 < \sum_{i=1}^N \mu_{ij} < N$ ، $\mu_{ij} \in [0, 1]$ و $\sum_{j=1}^C \mu_{ij} = 1$ وجود دارد که مقادیر متغیرهای i و j به ترتیب در بازه‌های $1 \leq i \leq N$ و $1 \leq j \leq C$ تغییر می‌کنند.

بر اساس تابع هدف معرفی‌شده در فرمول (۱)، معادلات به‌روزرسانی مراکز و توابع تعلق به‌صورت زیر خواهد بود:

$$\mu_{ij} = \frac{1}{\sum_{l=1}^C \left(\frac{d^2(x_i - v_j)}{d^2(x_i - v_l)} \right)^{\frac{1}{m-1}}} \quad (2)$$

$$v_j = \frac{\sum_{i=1}^N \mu_{ij}^m x_i}{\sum_{i=1}^N \mu_{ij}^m} \quad (3)$$

۳- الگوریتم پیشنهادی

در این بخش ابتدا الگوریتم پیشنهادی خوشه‌بندی فازی سی- مینز با وزن‌دهی اتوماتیک محلی (RLFWFCM^۳) فرمول‌بندی شده، سپس در قالب یک شبه‌کد با جزئیات کامل توضیح داده می‌شود. در ادامه اثبات همگرایی و پیچیدگی محاسباتی آن ذکر خواهد شد.

خوشه‌بندی از جمله رویکردهای مهم در داده‌کاوی، یادگیری ماشین و تشخیص الگو است. روش‌های خوشه‌بندی می‌تواند به دودسته سلسله‌مراتبی^۱ [۱-۳] و افرازکننده^۲ [۴-۷] تقسیم شود. الگوریتم‌های سلسله‌مراتبی از یک معیار شباهت استفاده کرده و در هر مرحله داده‌ها را به دودسته تقسیم و در نهایت یک ساختار درختی از این تقسیم‌بندی تحت عنوان دندروگرام^۲ ایجاد می‌کنند. الگوریتم‌های افرازکننده، داده‌ها را به‌طور مستقیم درون چند خوشه قرار می‌دهند. این الگوریتم‌ها خود به دودسته سخت^۴ (یا انحصاری^۵) [۸-۱۲] و نرم^۶ (یا فازی^۷) [۱۳-۱۶] تقسیم می‌شوند. در خوشه‌بندی سخت یک نمونه ورودی فقط به یک خوشه تعلق می‌گیرد در حالی که در خوشه‌بندی نرم، میزان تعلق آن به هر خوشه با یک عدد بین صفر و یک تعیین می‌شود.

الگوریتم‌های خوشه‌بندی هارد سی- مینز (HCM^۸) [۸-۱۲] و فازی سی- مینز (FCM^۹) [۱۳-۱۶] دو الگوریتم پایه از دسته الگوریتم‌های افرازکننده هستند. این دو الگوریتم به دلیل مزایایی که دارند همواره مورد توجه پژوهش‌گران می‌باشند. از جمله این مزایا می‌توان به ساختار ساده، پیاده‌سازی آسان، همگرایی سریع و نیاز به فضای ذخیره‌سازی کم - به دلیل ارائه هر خوشه توسط مرکز ثقل یا میانگین - اشاره کرد. در بسیاری از کاربردهای واقعی، تعداد ویژگی‌ها بسیار زیاد بوده که تعداد زیادی از این ویژگی‌ها غیر مرتبط هستند و نسبت به دیگر ویژگی‌ها ممکن است اهمیت کم‌تری داشته باشند. این در حالی است که در این الگوریتم‌ها برای تمام ویژگی‌ها اهمیت یکسان در نظر گرفته و این مسئله باعث کاهش کارایی آن‌ها می‌شود [۸، ۲۵-۱۷]. از طرف دیگر با توجه به این‌که داده‌های واقعی دارای نویز هستند، استفاده از فاصله اقلیدسی در این الگوریتم‌ها نیز عامل دیگری در جهت کاهش کارایی آن‌ها در کاربردهای واقعی است [۲۶، ۲۵].

برای رفع این مشکل، راهکارهای مختلفی به‌منظور انتخاب ویژگی و یا اختصاص وزن به ویژگی‌ها در فرآیند خوشه‌بندی ارائه شده است [۸، ۲۵-۱۷]. در تعدادی از این راهکارها، ابتدا وزن ویژگی‌ها تعیین شده و سپس خوشه‌بندی بر مبنای این اوزان انجام می‌شود. مشکل این الگوریتم‌ها این است که وزن ویژگی‌ها در تمام فرآیند خوشه‌بندی ثابت است [۱۸، ۲۱، ۲۳]. برای رفع این کاستی راهکارهای مختلفی برای تطبیق‌پذیر بودن وزن ویژگی‌ها در طول فرآیند خوشه‌بندی ارائه شده است [۱۷، ۲۱، ۲۲، ۲۴]. اخیراً نیز در سال ۲۰۱۴ ژیاو^{۱۰} و همکارانش [۲۵] یک الگوریتم خوشه‌بندی مبتنی بر هارد سی- مینز با وزن‌دهی ویژگی‌ها به‌صورت اتوماتیک و در طول فرآیند خوشه‌بندی ارائه دادند.

در این مقاله با بهره‌گیری از ایده مقاله [۲۵] یک روش جدید خوشه‌بندی فازی با وزن‌دهی محلی ویژگی‌ها ارائه شده است. در این روش هرکدام از ویژگی‌ها به‌صورت محلی وزن‌دهی می‌شوند، بدین ترتیب تأثیر هر ویژگی در خوشه‌های مختلف متفاوت خواهد بود. از طرفی وزن‌دهی ویژگی‌ها به‌صورت اتوماتیک و در زمان یادگیری و خوشه‌بندی انجام خواهد شد. همچنین در این الگوریتم با الهام گرفتن

$$w_{jk} = \begin{cases} \frac{1}{m_k} & \text{if } D_{jk} = 0, m_k = \{l : D_{jl} = 0\} \\ 0 & \text{if } D_{jk} \neq 0 \text{ but } \exists l \text{ s.t. } D_{jl} = 0 \\ \frac{1}{\sum_{l=1}^M \left(\frac{D_{jk}}{D_{jl}}\right)^{\beta-1}} & \text{if } D_{jl} \neq 0 \\ \forall 1 \leq l \leq M \end{cases} \quad (8)$$

که D_{jk} مجموع فواصل درون خوشه‌ای با توجه به صفت k ام و فاصله غیراقلیدسی است که به صورت زیر تعریف می‌شود:

$$D_{jk} = \sum_{i=1}^N \mu_{ij}^m (1 - \exp(-\gamma_k (x_{ik} - v_{jk})^2)) \quad (9)$$

جزئیات به دست آوردن معادلات (۶) تا (۸) در بخش ضمایم نشان داده شده است.

مراحل الگوریتم پیشنهادی در شبه‌کد ۱ خلاصه شده است. لازم به ذکر است که عناصر ماتریس اوزان در مرحله مقداردهی اولیه یکسان هستند، به طوری که $w_{jk} = \frac{1}{k}$ که $k=1, \dots, M$ که M تعداد ویژگی‌ها در مجموعه داده است. از طرفی مراکز اولیه به تصادف از بین نمونه‌ها انتخاب می‌شوند.

شبه‌کد ۱. الگوریتم خوشه‌بندی پیشنهادی

ورودی: مجموعه داده $D=[x_{ik}]$ به طوری که $i=1, \dots, N$ و $k=1, \dots, M$ تعداد نمونه‌ها و M تعداد صفات و x_{ik} صفت k ام از نمونه i ام است.

خروجی: ماتریس تعلق μ ، ماتریس مراکز V و ماتریس اوزان W .
مرحله ۱: مقداردهی اولیه تعداد خوشه‌های C ، درجه فازی‌سازی m ، مقدار آستانه ϵ در شرط توقف، تعداد تکرارهای T ، ماتریس اوزان W و ماتریس مراکز V .

مرحله ۲: به روزرسانی ماتریس تعلق μ از روی فرمول (۶).

مرحله ۳: به روزرسانی ماتریس مراکز V از روی فرمول (۷).

مرحله ۴: به روزرسانی ماتریس اوزان W از روی فرمول (۸).

مرحله ۵: اگر شرط $|v_j^i - v_j^{i-1}|^2 < \epsilon$ برقرار نیست به طوری که v_j^i مرکز خوشه j ام در تکرار i ام است و یا تعداد تکرارها کم‌تر از T است به مرحله ۲ برو. در غیر این صورت به مرحله ۶ برو.

مرحله ۶: گزارش ماتریس مراکز نهایی V ، ماتریس تعلق μ و ماتریس اوزان ویژگی‌های W .

۳-۲- اثبات همگرایی

در این بخش اثبات همگرایی الگوریتم پیشنهادی توضیح داده می‌شود. در واقع هدف، اثبات این موضوع است که تابع موجود در رابطه (۵) با توجه به فرمول‌هایی که برای به روزرسانی اجزای آن به دست آمده در نهایت به مقدار کمینه خود می‌رسد یا به عبارتی همگرا می‌شود. با توجه به اینکه تابع هدف، مجموع فواصل درون خوشه‌ای است هرچه مقدار آن کم‌تر باشد بهتر است.

قضیه ۱. فرض کنید ماتریس مراکز V و ماتریس اوزان W ثابت باشد، آنگاه ماتریس تعلق μ یک کمینه محلی از رابطه (۵) است اگر و

۳-۱- فرمول‌بندی الگوریتم پیشنهادی

برای خوشه‌بندی نیاز به یک تابع هدف است که بر اساس آن پارامترهای الگوریتم خوشه‌بندی تنظیم می‌شوند. یکی از توابع معروف در این زمینه، مجموع فواصل درون خوشه‌ای است. از طرفی در بسیاری از مقالات از معیار شباهت اقلیدسی برای محاسبه تابع هدف استفاده شده است اما این معیار تا حد زیادی به نویز حساس است. همچنین در این معیارها معمولاً همه ویژگی‌ها با وزن یکسان در نظر گرفته می‌شوند.

در الگوریتم پیشنهادی از یک تابع هدف جدید که از معیار شباهت غیر اقلیدسی که در مقاله [۲۶] ارائه شده است، استفاده می‌شود. همچنین در این تابع هدف، میزان تعلق یک نمونه به یک خوشه یک عدد بین صفر و یک است و به ازای هر ویژگی در هر خوشه وزن مختص به آن ویژگی در نظر گرفته می‌شود. تابع هدف الگوریتم پیشنهادی به صورت زیر است:

$$J = \sum_{i=1}^N \sum_{j=1}^C \sum_{k=1}^M \mu_{ij}^m w_{jk}^\beta d^2(x_{ik} - v_{jk}) \quad (4)$$

در اینجا μ_{ij} میزان تعلق نمونه i ام به مرکز خوشه j ام، w_{jk} وزن ویژگی k ام در خوشه j ام و $d^2(x_{ik} - v_{jk})$ از نوع فاصله غیراقلیدسی است و برابر با $1 - \exp(-\gamma_k (x_{ik} - v_{jk})^2)$ تعریف می‌شود [۲۶]. برای این تابع هدف، نیز محدودیت‌های $0 < \sum_{i=1}^N \mu_{ij} < N$ ، $\mu_{ij} \in [0, 1]$ ، $\sum_{j=1}^C \mu_{ij} = 1$ و $w_{jk} \in [0, 1]$ ، $\sum_{k=1}^M w_{jk} = 1$ وجود دارد که مقادیر متغیرهای i ، j و k به ترتیب در بازه‌های $1 \leq i \leq N$ ، $1 \leq j \leq C$ و $1 \leq k \leq M$ تغییر می‌کنند.

در اینجا C تعداد خوشه‌ها و M تعداد ویژگی‌هاست. با توجه به محدودیت‌هایی که در بالا اشاره شد می‌توان معادله لاگرانژ تابع هدف الگوریتم پیشنهادی را به صورت زیر نوشت:

$$\tilde{J} = \sum_{i=1}^N \sum_{j=1}^C \sum_{k=1}^M \mu_{ij}^m w_{jk}^\beta d^2(x_{ik} - v_{jk}) - \lambda \left(\sum_{j=1}^C \mu_{ij} - 1 \right) - \alpha \left(\sum_{k=1}^M w_{jk} - 1 \right) \quad (5)$$

که λ و α پارامترهای معادله لاگرانژ هستند.

با حل معادله لاگرانژ فوق با توجه به شرایط KKT* فرمول‌های به روزرسانی زیر برای μ_{ij} ، v_{jk} و w_{jk} به دست می‌آیند. در واقع با مشتق‌گیری از تابع هدف فوق نسبت به v_{jk} و w_{jk} و مساوی صفر قرار دادن مشتق و حل معادله به دست آمده فرمول‌های زیر حاصل می‌شوند:

$$\mu_{ij} = \frac{1}{\sum_{l=1}^C \left(\sum_{k=1}^M w_{jk}^\beta d^2(x_{ik} - v_{jk}) \right)^{\frac{1}{m-1}} + \sum_{k=1}^M w_{lk}^\beta d^2(x_{ik} - v_{lk})} \quad (6)$$

$$v_{jk} = \frac{\sum_{i=1}^N \mu_{ij}^m \exp(-\gamma_k (x_{ik} - v_{jk})^2) x_{ik}}{\sum_{i=1}^N \mu_{ij}^m \exp(-\gamma_k (x_{ik} - v_{jk})^2)} \quad (7)$$

۳-۳- پیچیدگی محاسباتی

با توجه به شبه‌کد الگوریتم پیشنهادی می‌توان دریافت که پیچیدگی محاسباتی مرحله دوم معادل NMC^2 است که N تعداد نمونه‌ها، C تعداد خوشه‌ها و M تعداد ویژگی‌ها است. در مرحله دوم با دو مجموع تودرتو مواجهیم که اولی C بار و دومی M بار تکرار می‌شود و این عمل به ازای هر نمونه در هر خوشه انجام می‌شود. از طرفی پیچیدگی محاسباتی مرحله سوم معادل NMC است چون با یک مجموع که N بار تکرار می‌شود مواجهیم و از طرفی این عمل به ازای هر ویژگی در هر خوشه انجام خواهد شد. در نتیجه پیچیدگی محاسباتی کل برابر است با $T(NC^2 + NMC) = TNCM(C + 1)$ که تقریباً برابر است با $O(TNC^2M)$. که T تعداد تکرارهای الگوریتم است. با این که پیچیدگی محاسباتی الگوریتم پیشنهادی زیاد است اما در بخش آزمایش‌ها نشان داده می‌شود که کارایی آن بالاست.

۴- نتایج آزمایش‌ها

در این بخش کارایی الگوریتم پیشنهادی (RLFWFCM) مورد بررسی قرار گرفته و نتایج آن با الگوریتم‌های خوشه‌بندی فازی سی-مینز با وزن‌دهی سراسری (RFWFCM) [۱۹] و هارد سی-مینز با وزن‌دهی اتوماتیک محلی (RLFWHCM) [۲۵] بر یک مجموعه داده مصنوعی و پنج مجموعه داده واقعی از انبار داده UCI مورد مقایسه قرار می‌گیرد. پارامترهای مشترک موجود در همه الگوریتم‌ها شامل مقدار آستانه ϵ و تعداد تکرارهای T به صورت $\epsilon=10^{-5}$ و $T=200$ تنظیم شده است. همچنین پارامتر m در الگوریتم RFWFCM و الگوریتم پیشنهادی به صورت $m=2$ تنظیم شده است. تعداد خوشه‌ها نیز به ازای هر مجموعه داده همان تعداد کلاس‌های موجود در مجموعه داده مورد استفاده قرار داده شد. در این میان از مقدار تابع هدف و نرخ خطا برای اندازه‌گیری کارایی الگوریتم‌ها استفاده می‌شود.

۴-۱- مجموعه داده مصنوعی

برای آزمایش کارایی الگوریتم پیشنهادی یک مجموعه داده مصنوعی تولید شده است. توزیع نمونه‌ها در این مجموعه داده از نوع گوسین است اما برای خوشه‌های مختلف مقادیر میانگین‌ها و انحراف از معیار متفاوت است. در این مجموعه داده ۴۰۰ نمونه چهاربعدی در ۴ کلاس با ۱۰۰ نمونه در هر کلاس وجود دارد. میانگین‌ها و انحراف از معیارها به ترتیب به صورت زیر است:

$$\begin{aligned}\mu_{C1} &= [0.5, -0.5, 0, 0] \\ \mu_{C2} &= [-0.5, -0.5, 0, 0] \\ \mu_{C3} &= [0, 0.5, 0.5, 0] \\ \mu_{C4} &= [0, 0.5, -0.5, 0] \\ \sigma_{C1} = \sigma_{C2} &= [0, 2, 0, 2, 0, 6, 0, 6] \\ \sigma_{C3} = \sigma_{C4} &= [0, 6, 0, 2, 0, 2, 0, 6]\end{aligned}$$

فقط اگر هر عنصر از آن یعنی توسط رابطه (۶) تعریف شود. چون تابع هدف با ماتریس تعلق رابطه مستقیم دارد.

اثبات. همان‌طور که مراحل در بخش ضمایم موجود است μ_{ij} از مشتق گرفتن از رابطه (۵) نسبت به μ_{ij} و مساوی صفر قرار دادن آن به دست آمده است؛ در نتیجه یا کمینه است یا بیشینه. حال اگر یکبار دیگر از رابطه (۵) نسبت به μ_{ij} مشتق بگیریم و نشان دهیم که مشتق جزئی دوم رابطه (۵) نسبت به μ_{ij} مثبت است می‌توان ثابت کرد که μ_{ij} تعریف شده توسط رابطه (۶) یک کمینه محلی از رابطه (۵) است. بنابراین در ادامه از رابطه (۵) نسبت به μ_{ij} مشتق دوم می‌گیریم. حاصل این مشتق به صورت زیر خواهد بود:

$$\sum_{k=1}^M m(m-1)\mu_{ij}^{m-2} w_{jk}^\beta d^2(x_{ik} - v_{jk}) \quad (10)$$

از آنجا که $d^2(x_{ik} - v_{jk}) > 0$ و $w_{jk} > 0$ و $m > 1$ می‌توان نشان داد $\sum_{k=1}^M m(m-1)\mu_{ij}^{m-2} w_{jk}^\beta d^2(x_{ik} - v_{jk}) > 0$. در نتیجه رابطه (۶) یک کمینه محلی از رابطه (۵) است.

قضیه ۲. فرض کنید ماتریس تعلق μ و ماتریس اوزان W ثابت باشد. آنگاه ماتریس مراکز W یک کمینه محلی از رابطه (۵) است اگر و فقط اگر v_{jk} توسط رابطه (۷) تعریف شود.

اثبات. اثبات قضیه ۲ نیز مانند اثبات قضیه ۱ است. یعنی تنها کافی است که نشان دهیم مشتق دوم رابطه (۵) نسبت به v_{jk} مثبت است چون v_{jk} از طریق مشتق گرفتن از رابطه (۵) نسبت به v_{jk} و مساوی صفر قرار دادن آن به دست آمده است. مشتق جزئی دوم رابطه (۵) نسبت به v_{jk} یک عبارت مثبت است. در نتیجه رابطه (۷) نیز یک کمینه محلی از رابطه (۵) است.

قضیه ۳. فرض کنید ماتریس تعلق μ و ماتریس مراکز V ثابت باشد. آنگاه ماتریس اوزان W یک کمینه محلی از رابطه (۵) است اگر و فقط اگر w_{jk} توسط رابطه (۸) تعریف شود.

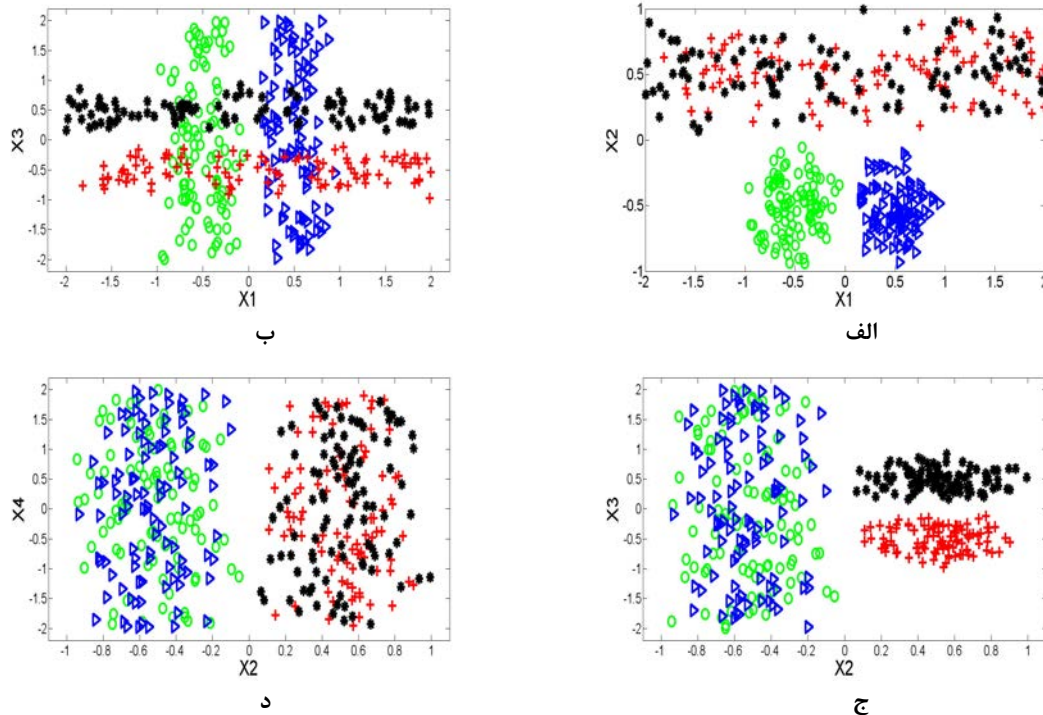
اثبات. اثبات قضیه ۳ نیز مانند اثبات قضیه ۱ و ۲ است. یعنی تنها کافی است که نشان دهیم مشتق دوم رابطه (۵) نسبت به w_{jk} مثبت است چون w_{jk} از طریق مشتق گرفتن از رابطه (۵) نسبت به w_{jk} و مساوی صفر قرار دادن آن به دست آمده است. مشتق جزئی دوم رابطه (۵) نسبت به w_{jk} برابر است با $\sum_{i=1}^N \mu_{ij}^m \beta w_{jk}^{\beta-1} d^2(x_{ik} - v_{jk})$ که یک عبارت مثبت است. در نتیجه رابطه (۸) نیز یک کمینه محلی از رابطه (۵) است.

از آنجا که هدف، کمینه کردن رابطه (۵) است، لذا با توجه به قضیه‌های ۱ تا ۳ و همچنین با توجه به فرمول‌های به‌روزرسانی ماتریس مراکز V (رابطه ۶) و ماتریس اوزان W (رابطه ۷) و ماتریس تعلق μ (رابطه ۸)، می‌توان اطمینان داشت که تابع هدف در نهایت به مقدار کمینه خود می‌رسد یا همگرا می‌شود.

در جدول ۱ مشخصات این مجموعه داده ارائه شده است. شکل ۱ نیز مجموعه داده مصنوعی را در زیرفضاهای مختلف از ویژگی‌ها نشان می‌دهد.

جدول ۱: مشخصات مجموعه داده مصنوعی

تعداد مشخصه	نام مشخصه
۴۰۰	نمونه
۴	بعد (ویژگی)
۴	کلاس



شکل ۱: مجموعه داده مصنوعی در زیرفضاهای مختلف ویژگی‌ها: الف) $X1$ و $X2$ ، ب) $X2$ و $X3$ ، ج) $X3$ و $X1$ ، د) $X4$ و $X2$

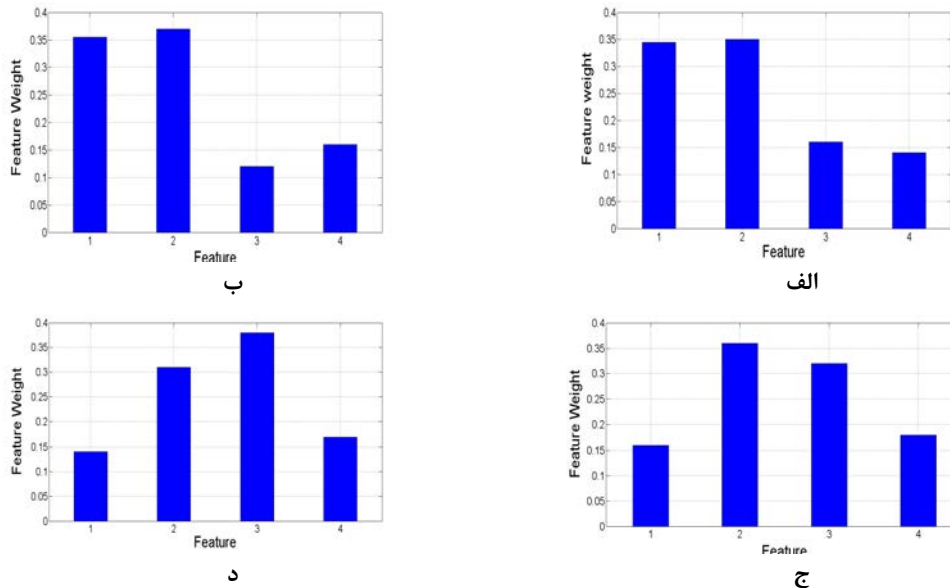
لازم به ذکر است که همه این الگوریتم‌ها به مراکز اولیه حساس‌اند و نتیجه یک اجرا نمی‌تواند اثبات‌کننده خوب یا بد بودن یک الگوریتم باشد. لذا نتایج ذکر شده شامل نمودارهای همگرایی و جداول مربوطه نتیجه میانگین ۱۰ اجراست. از طرفی در مقایسه الگوریتم‌ها بر هر مجموعه داده از یک سری مراکز اولیه و پارامترهای یکسان استفاده شده است چون تنها در صورت شرایط یکسان می‌توان چند الگوریتم را از هر لحاظی باهم مقایسه کرد. جدول ۲ میانگین نرخ خطا و زمان اجرا و تابع هدف دو الگوریتم RFWFCM و RLFWHCM و الگوریتم پیشنهادی بعد از ۲۰۰ بار تکرار بر مجموعه داده مصنوعی به ازای $\beta=2$ را نشان می‌دهد.

با توجه به جدول ۲ می‌توان دریافت که زمان اجرای الگوریتم پیشنهادی از زمان اجرای الگوریتم RLFWHCM بیش‌تر است. البته این خود امری کاملاً صحیح است چون محاسبه میزان تعلق هر نمونه به هر خوشه زمان‌بر است. از طرفی زمان اجرای الگوریتم پیشنهادی حتی از زمان اجرای RFWFCM نیز بیش‌تر است و دلیل آن می‌تواند وزن‌دهی محلی در الگوریتم پیشنهادی باشد. با وجود تمام این موارد تابع هدف و نرخ خطای الگوریتم پیشنهادی به ترتیب از تابع هدف و

از روی شکل ۱ مخصوصاً قسمت‌های الف و ب می‌توان دریافت داده‌های خوشه‌های ۱ و ۲ (نقاط با رنگ سبز (علامت دایره) و رنگ آبی (علامت مثلث)) به کمک ویژگی‌های ۱ و ۲ و داده‌های خوشه‌های ۳ و ۴ (نقاط با رنگ مشکی (علامت *)) و رنگ قرمز (علامت +)) به کمک ویژگی‌های ۲ و ۳ خوشه‌بندی می‌شوند. یعنی باید وزن ویژگی‌های ۱ و ۲ در خوشه‌های ۱ و ۲ و وزن ویژگی‌های ۲ و ۳ در خوشه‌های ۳ و ۴ بیش از دیگر ویژگی‌ها در این خوشه‌ها باشد. حال باید دید آیا اجرای الگوریتم پیشنهادی بر این مجموعه داده مصنوعی حقیقت فوق را نشان می‌دهد یا خیر؟

شکل ۲ الف نشان می‌دهد که ویژگی‌های ۱ و ۲ در خوشه اول وزن بیش‌تری نسبت به سایر ویژگی‌ها کسب کرده‌اند. از طرفی با مشاهده شکل ۱ الف می‌توان مشاهده کرد که ویژگی‌های ۱ و ۲ تأثیر زیادی در خوشه‌بندی خوشه ۱ دارند. وزن‌های به‌دست‌آمده در شکل‌های ۲ ج و ۲ د نیز نشان‌دهنده تأثیر متفاوت ویژگی‌ها در خوشه‌های مختلف است. بدین ترتیب می‌توان مطمئن بود که الگوریتم پیشنهادی با وجود یک مجموعه داده شبیه شکل ۱ دارای عملکرد صحیحی است و وزن ویژگی‌ها را به‌درستی تعیین می‌کند.

نرخ خطای هر دو الگوریتم دیگر کمتر است و این خود دلیل مناسبی بر افزایش کارایی آن نسبت به دو الگوریتم دیگر است.



شکل ۲: وزن ویژگی‌ها در خوشه‌ها؛ (الف) اول، (ب) دوم، (ج) سوم، (د) چهارم، با اجرای الگوریتم پیشنهادی بر روی مجموعه داده مصنوعی

جدول ۲: میانگین زمان اجرا و تابع هدف و نرخ خطای الگوریتم‌های RFWFCM، RLFWHCM و RLFWFCM بعد از ۲۰۰ بار تکرار بر مجموعه داده مصنوعی

روش	زمان (میلی ثانیه)	تابع هدف		نرخ خطا	
		میانگین	انحراف از معیار	میانگین	انحراف از معیار
RFWFCM	۱۶۴۵۶/۲	۹/۸۳	۱/۱۷	۲۱/۵۷	۲/۶
RLFWHCM	۷۰۵	۱۳/۶۷	۲/۵۹	۲۳/۷۷	۲/۴۴
RLFWFCM	۱۶۴۸۱	۷/۳۶	۰/۴۱	۱۲/۴۵	۲/۰۶

جدول ۳: میانگین تابع هدف RFWFCM، RLFWHCM و RLFWFCM بر مجموعه داده مصنوعی به ازای مقادیر مختلف β

روش	β					
	۶	۴	۲	۲	۲	۶
RFWFCM	۰/۰۴۲	۰/۱۶۶	۹/۸۳	۲۷۵۰/۲۸	۴۳۹۶۰/۱۴	۷۰۳۰۳۱
RLFWHCM	۰/۰۷۹	۱/۱۸	۱۳/۶۷	۶۲۱۱/۴۱	۹۲۳۸۹/۷۸	۱۴۳۷۳۱۴
RLFWFCM	۰/۰۴۰	۰/۱۶۳	۷/۳۶	۲۷۱۸/۶	۴۳۳۳۰/۶۷	۶۹۲۰۱۶

با توجه به مقادیر به دست آمده از روی جدول ۳ و ۴ می‌توان نتیجه گرفت الگوریتم پیشنهادی به ازای تمام مقادیر β از سایر الگوریتم‌ها بهتر است چون تابع هدف و نرخ خطای آن کمتر است. از روی جدول ۴ می‌توان نتیجه گرفت نرخ خطای الگوریتم RFWFCM به ازای $\beta=2$ و نرخ خطای الگوریتم RLFWHCM به ازای $\beta=6$ و نرخ خطای الگوریتم RLFWFCM به ازای $\beta=2$ از نرخ خطا به ازای سایر مقادیر β کمتر است. این نتیجه‌گیری از روی جدول ۳ قابل استنتاج نیست چون همان‌طور که از روی معادله (۴) قابل برداشت است هرچه β بیشتر باشد مقدار تابع هدف کم‌تر خواهد بود؛ چون ماتریس اوزان مقادیری بین صفر و یک را اخذ می‌کند.

جدول ۳ میانگین تابع هدف RFWFCM، RLFWHCM و الگوریتم پیشنهادی بعد از ۲۰۰ بار تکرار بر مجموعه داده مصنوعی استفاده شده به ازای مقادیر مختلف β ، و جدول ۴ میانگین نرخ خطای RFWFCM، RLFWHCM و الگوریتم پیشنهادی بعد از ۲۰۰ بار تکرار بر مجموعه داده مصنوعی استفاده شده به ازای مقادیر مختلف β را نشان می‌دهد.

جدول ۴: مقایسه میانگین نرخ خطا بر مجموعه داده مصنوعی به ازای مقادیر مختلف β

β						روش
-۶	-۴	-۲	۲	۴	۶	
۱۷/۶۲	۱۷/۷۲	۲۱/۵۷	۱۷/۱۲	۱۷/۳۲	۱۷/۵۷	RFWFCM
۱۸/۱۷	۱۹/۸	۲۳/۷۷	۲۰/۰۷	۱۸/۹	۱۷/۲۷	RLFWHCM
۱۶/۳۲	۱۶/۶۷	۱۲/۴۵	۱۶/۵۷	۱۷/۲۲	۱۷/۰۵	RLFWFCM

پنج مجموعه داده استفاده شده به ازای مقادیر مختلف β را نشان می‌دهد.

در این مقاله پارامتر β پارامتری مهم و بر کارایی الگوریتم‌ها تأثیرگذار است. در اینجا با در نظر گرفتن مقادیر صحیح بین ۶- و ۶+ برای پارامتر β تأثیر آن بر کارایی الگوریتم‌ها مورد مطالعه قرار می‌گیرد. بدین منظور به ازای هر مجموعه داده ۱۰ سری مرکز تولید شده و هر الگوریتم ۱۰ بار بر هر مجموعه داده و هر بار طی ۲۰۰ تکرار اجرا می‌شود و میانگین تابع هدف الگوریتم‌ها در جدول ۷ و میانگین نرخ خطای آن‌ها در جدول ۸ به نمایش درمی‌آید. از طرفی شکل ۳ نیز نتایج موجود در جدول ۸ را به تصویر می‌کشد. در شکل ۳ محور عمودی نرخ خطا و محور افقی مقدار پارامتر β است.

از روی جدول ۷ می‌توان دریافت که مقدار تابع هدف الگوریتم پیشنهادی برای تمام مجموعه داده‌های مورد استفاده بعد از ۲۰۰ تکرار به ازای مقادیر مختلف β از دو الگوریتم دیگر کم‌تر است. اما از روی جدول ۸ می‌توان دریافت که نرخ خطای الگوریتم پیشنهادی برای تمام مجموعه داده‌ها به ازای تمام مقادیر مختلف β از نرخ خطای دو الگوریتم دیگر کم‌تر نیست. در واقع برای مجموعه داده glass به ازای $\beta=4$ ، برای bupa به ازای $\beta=-6$ و برای breast به ازای $\beta=-2$ نرخ خطای الگوریتم پیشنهادی از نرخ خطای دو الگوریتم دیگر کم‌تر نیست در نتیجه تنها در این موارد که البته تعداد آن‌ها اندک است الگوریتم پیشنهادی نسبت به دو الگوریتم دیگر کارا تر نیست. از روی جدول ۷ و جدول ۸ و شکل ۳ می‌توان نشان داد که الگوریتم پیشنهادی به ازای اکثر مقادیر مختلف پارامتر β از سایر الگوریتم‌های مقایسه شده بهتر است. علاوه بر این از روی شکل ۳ به راحتی می‌توان مقدار مناسب پارامتر β را برای هر مجموعه داده تعیین کرد. مثلاً این که مقدار مناسب پارامتر β برای الگوریتم پیشنهادی به ازای مجموعه داده iris معادل ۲، به ازای مجموعه داده wine معادل ۶، به ازای مجموعه داده glass معادل ۶، به ازای مجموعه داده bupa معادل ۲، و به ازای مجموعه داده breast-cancer نیز معادل ۲ است.

۴-۲- مجموعه داده‌های واقعی

در این بخش یک ارزیابی از الگوریتم پیشنهادی در مقایسه با دو روش RFWFCM و RLFWHCM بر پنج مجموعه داده واقعی شامل iris، wine، glass، bupa و breast-cancer از انبار داده UCI انجام می‌شود. اطلاعات مربوط به این مجموعه داده‌ها در جدول ۵ خلاصه شده است.

جدول ۵: مشخصات پنج مجموعه داده واقعی

نام	تعداد کلاس‌ها	تعداد ابعاد	تعداد نمونه‌ها
Iris	۳	۴	۱۵۰
Wine	۳	۱۳	۱۷۸
Glass	۶	۱۰	۲۱۴
Bupa	۲	۶	۳۴۵
Breast	۲	۱۰	۶۸۳

جدول ۶ میانگین نرخ خطا و زمان اجرا و تابع هدف الگوریتم پیشنهادی بعد از ۲۰۰ بار تکرار بر پنج مجموعه داده استفاده شده به ازای $\beta=2$ را نشان می‌دهد. با توجه به جدول ۶ می‌توان دریافت که زمان اجرای الگوریتم پیشنهادی از زمان اجرای الگوریتم RLFWHCM به ازای تمام مجموعه داده‌ها بیشتر است. از طرفی زمان اجرای الگوریتم پیشنهادی حتی از زمان اجرای RFWFCM نیز بیشتر است دلیل بیشتر بودن زمان اجرا به ازای این مجموعه داده‌ها مشابه دلیل بیشتر بودن زمان اجرا به ازای مجموعه داده مصنوعی است. با وجود تمام این‌ها به ازای تمام مجموعه داده‌های مورد استفاده تابع هدف و نرخ خطای الگوریتم پیشنهادی به ترتیب از تابع هدف و نرخ خطای هر دو الگوریتم دیگر کم‌تر است و این خود دلیل محکمی بر افزایش کارایی آن نسبت به دو الگوریتم دیگر بعد از اجرا بر مجموعه داده‌های واقعی است.

جدول ۷ میانگین نرخ خطای الگوریتم RFWFCM، RLFWHCM و RFWFCM بعد از ۲۰۰ بار تکرار بر پنج مجموعه داده استفاده شده به ازای مقادیر مختلف β و جدول ۸ میانگین تابع هدف الگوریتم RFWFCM، RLFWHCM و RLFWHCM بعد از ۲۰۰ بار تکرار بر

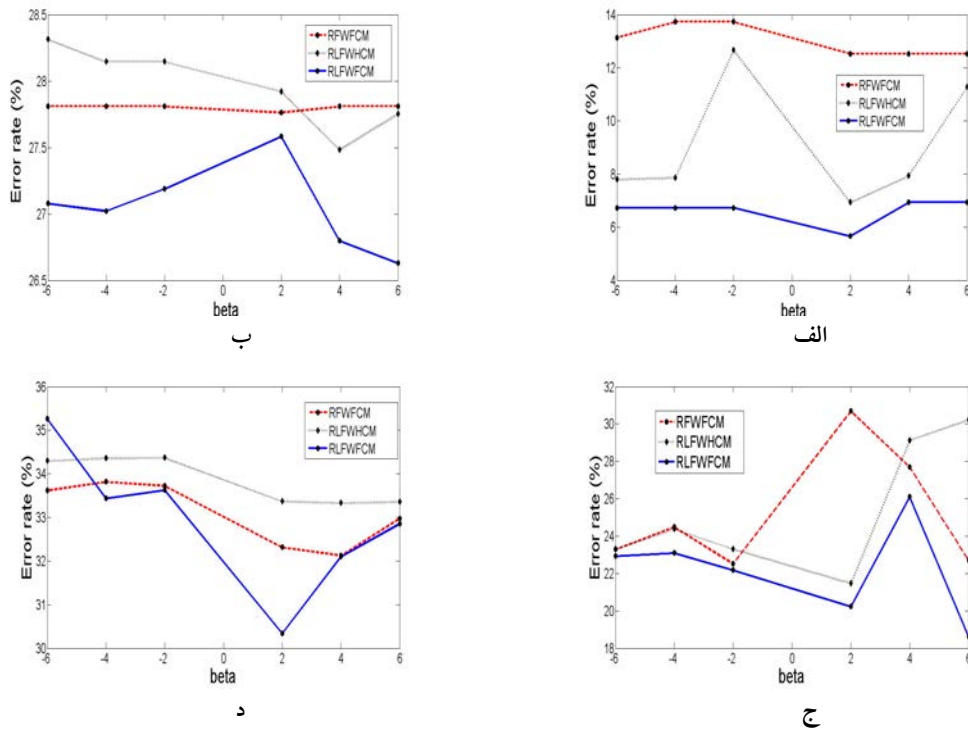
جدول ۶: زمان اجرا و تابع هدف و میانگین نرخ خطای الگوریتم‌های RLFWFCM و RLFWHCM

نرخ خطا		تابع هدف		زمان (میلی ثانیه)		روش	مجموعه داده
انحراف از معیار	میانگین	انحراف از معیار	میانگین	میانگین	میانگین		
۴/۳۳	۱۲/۱۵۳	۰/۴۴	۲/۵۹	۸۲۴۲/۶		RLFWFCM	iris
۰/۶۸	۶/۹۳	۰/۰۰۷	۳/۲۱	۱۹۹		RLFWHCM	
۰/۰۰۰۲	۵/۶۶	۰/۰۰۳	۲/۳۶	۸۲۶۹/۱		RLFWFCM	
۰/۰۰۰۱	۲۷/۷۶	۰/۰۰۰۱	۱/۸۲	۱۲۱۱۶		RLFWFCM	wine
۱/۲۲	۲۷/۹۲	۰/۳	۳/۱۱	۴۶۵/۷		RLFWHCM	
۰/۰۰۲	۲۷/۵۸	۰/۰۰۰۳	۱/۷۵	۱۲۱۳۰/۱		RLFWFCM	
۲/۶۳	۳۰/۳۷۰	۰/۰۵	۰/۸۲	۱۳۲۵۶/۶		RLFWFCM	glass
۱/۰۴	۲۱/۴۷	۰/۱۱	۰/۱	۱۳۱۷/۹		RLFWHCM	
۰/۴۲	۲۰/۲۳	۰/۰۰۶	۰/۰۰۱	۱۳۲۹۸/۵		RLFWFCM	
۳/۷۳	۳۲/۳۲	۰/۰۰۰۲	۸/۶۷	۱۲۹۷۷/۳		RLFWFCM	bupa
۰/۷۴	۳۳/۳۷	۰/۰۱	۱۱/۹۸	۶۷۹/۴		RLFWHCM	
۴/۴	۳۰/۳۴	۰/۰۰۰۱	۸/۶۲	۱۲۹۸۵/۱		RLFWFCM	
۳/۰۹	۳۷/۵۴	۰/۱۹	۸/۰۲	۵۰۷۰۲/۷		RLFWFCM	breast
۳/۴۶	۳۷/۴۲	۱/۲۴	۵/۱۷	۱۴۴۲/۷		RLFWHCM	
۲/۰۳	۳۱/۷۸	۰/۰۲	۴/۴۱	۵۰۷۰۷/۲		RLFWFCM	

جدول ۷: میانگین تابع هدف RLFWFCM و RLFWHCM بعد از ۲۰۰ بار تکرار بر پنج مجموعه داده واقعی استفاده شده به ازای

مقادیر مختلف β

β						روش	مجموعه داده
-۶	-۴	-۲	۲	۴	۶		
۲۱۳۷۰/۱/۳	۱۳۴۵۹/۰/۴	۸۵۴/۷۵	۲/۵۹	۰/۱۸	۰/۰۱۱۹	RLFWFCM	iris
۲۹۶۳۴۳/۳	۱۸۸۹۱/۶۹	۱۲۸۹/۷۶	۳/۲۱	۰/۲۴	۰/۰۱۶۷	RLFWHCM	
۲۰۸۰۹۸/۶	۱۳۱۳۶/۶۲	۸۳۸/۵۹	۲/۳۶	۰/۱۷	۰/۰۱۱۶	RLFWFCM	
۱۵۰×۱۰^۷	۸۸۱۲۵۶۰	۵۲۵۶۹/۶۰	۱/۸۲	۰/۰۱۰۸	۰/۰۰۰۱	RLFWFCM	wine
۲۷۹×۱۰^۷	۱۶۵۸۱۱۰۰	۹۹۱۴۷/۴۴	۳/۱۱	۰/۰۱۹۴	۰/۰۰۰۱	RLFWHCM	
۱۴۸×۱۰^۷	۸۷۵۹۸۵۸	۵۱۹۸۳/۵۳	۱/۷۵	۰/۰۱۰۶	۰/۰۰۰۱	RLFWFCM	
۹۵۲۵۵۸۵۶	۹۵۶۱۸۹	۹۶۴۲/۴۳	۰/۸۲	۰/۰۰۹	۰/۰۰۰۱	RLFWFCM	glass
NaN	NaN	NaN	۰/۱	۰/۰۰۲	۰/۰۰۰۳	RLFWHCM	
۸۸۵۴۵۷۹۸	۸۹۶۳۸۴	۹۱۸۸/۷۶	۰/۰۰۱	.	.	RLFWFCM	
۱۵۲۸۶۶۵۹	۴۲۵۳۳۰	۱۱۸۵۸/۶۴	۸/۶۷	۰/۲۵	۰/۰۰۶۹	RLFWFCM	bupa
۲۵۲۳۴۳۳۸	۷۰۵۵۰۹	۱۹۸۵۴/۰۷	۱۱/۹۸	۰/۳۸	۰/۰۱۰۹	RLFWHCM	
۱۵۱۴۱۴۰۷	۴۲۱۲۹۵	۱۱۷۴۸/۲۰	۸/۶۲	۰/۲۴	۰/۰۰۶۸	RLFWFCM	
۸۵۸×۱۰^۶	۸۶۰۰۳۸۰	۸۶۵۸۸/۰۹	۸/۰۲	۰/۰۸	۰/۰۰۰۸	RLFWFCM	breast
۱۲۷۰×۱۰^۶	۱۲۲۴۱۹۷۱	۱۲۴۱۷۹/۵۷	۵/۱۷	۰/۱۵	۰/۰۰۱۷	RLFWHCM	
۸۰۶×۱۰^۶	۸۱۰۷۰۶۹	۸۲۵۰۳/۸۵	۴/۴۱	۰/۰۷	۰/۰۰۰۵	RLFWFCM	



شکل ۳: نمودار میانگین نرخ خطای RFWFCM، RLFWHCM و RLFWFCM بعد از ۲۰۰ بار تکرار بر مجموعه داده‌های واقعی به‌ازای مقادیر مختلف β : الف) iris، ب) wine، ج) glass، د) bupa

جدول ۸: میانگین نرخ خطای RFWFCM و RLFWHCM بعد از ۲۰۰ بار تکرار بر پنج مجموعه داده واقعی استفاده شده به ازای مقادیر مختلف β

β						روش	مجموعه داده
-۶	-۴	-۲	۲	۴	۶		
۱۳/۱۳	۱۳/۷۳	۱۳/۷۳	۱۲/۵۳	۱۲/۵۳	۱۲/۵۳	RFWFCM	iris
۷/۸۰	۷/۸۶	۱۲/۶۶	۶/۹۳	۷/۹۳	۱۱/۲۶	RLFWHCM	
۶/۷۳	۶/۷۳	۶/۷۳	۵/۶۶	۶/۹۳	۶/۹۳	RLFWFCM	
۲۷/۸۰	۲۷/۸۹	۲۷/۸۰	۲۷/۷۶	۲۷/۸۰	۲۷/۸۰	RFWFCM	Wine
۲۸/۳۱	۲۸/۱۰	۲۸/۱۴	۲۷/۹۲	۲۷/۴۸	۲۷/۷۵	RLFWHCM	
۲۷/۰۷	۲۷/۰۲	۲۷/۱۹	۲۷/۵۸	۲۶/۷۹	۲۶/۶۲	RLFWFCM	
۲۳/۲۹	۲۴/۴۹	۲۲/۵۳	۳۰/۷۰	۲۷/۷۱	۲۲/۷۱	RFWFCM	glass
۲۳/۲۹	۲۴/۳۸	۲۳/۳۱	۲۱/۴۷	۲۹/۱۲	۳۰/۳۳	RLFWHCM	
۲۲/۹۳	۲۳/۱۰	۲۲/۱۸	۲۰/۲۳	۲۹/۱۲	۱۸/۶۴	RLFWFCM	
۳۳/۶۲	۳۳/۸۲	۳۳/۷۳	۳۲/۳۲	۳۲/۱۳	۳۲/۹۸	RFWFCM	Bupa
۳۴/۳۰	۳۴/۳۶	۳۴/۳۷	۳۳/۳۷	۳۳/۳۳	۳۳/۳۶	RLFWHCM	
۳۵/۲۷	۳۳/۴۴	۳۳/۶۳	۳۰/۳۴	۳۲/۱۰	۳۲/۸۶	RLFWFCM	
۳۲/۲۵	۳۲/۱۲	۳۲/۲۵	۳۷/۵۴	۳۴/۰۲	۳۴/۰۴	RFWFCM	breast
۳۳/۷۰	۳۳/۳۵	۳۳/۷۹	۳۷/۴۲	۳۷/۸۱	۳۷/۸۱	RLFWHCM	
۳۲/۱۷	۳۱/۸۲	۳۲/۳۸	۳۱/۷۸	۳۳/۴۹	۳۲/۸۸	RLFWFCM	

۵- نتیجه‌گیری

تاکنون تحقیقات زیادی به‌منظور طراحی یک الگوریتم خوشه‌بندی با ویژگی‌های وزن‌دهی شده صورت گرفته است. اما در هیچیک از این‌ها به‌طور همزمان دو خاصیت فازی بودن و وزن‌دهی محلی اعمال نشده است علاوه بر اینکه در اغلب آن‌ها از معیار شباهت اقلیدسی که نسبت به نویز بسیار حساس است استفاده شده است. در این مقاله الگوریتم پیشنهادی با تعریف یک تابع هدف جدید و به دنبال آن تعریف فرمول‌های به‌روزرسانی پیشنهاد داده می‌شود. در این مقاله اثبات همگرایی و پیچیدگی محاسباتی الگوریتم پیشنهادی انجام شده است. نتایج آزمایش‌ها نشان می‌دهد که در کل کارایی الگوریتم پیشنهادی با توجه به کاهش تابع هدف و کاهش نرخ خطا در الگوریتم‌های مشابه، از آن‌ها بیش‌تر است.

یکی از کارهایی که می‌توان در آینده برای بهبود کارایی الگوریتم پیشنهادی انجام داد تعیین اتوماتیک تعداد خوشه‌ها است، که در این میان می‌توان از استراتژی‌های مبتنی بر اعتبار خوشه [۲۸، ۲۹] استفاده کرد. کار بعدی که می‌توان انجام داد توسعه الگوریتم پیشنهادی برای اجرا بر مجموعه داده‌های اسمی است.

پیوست‌ها

در اینجا ابتدا مراحل به دست آوردن فرمول به‌روزرسانی تابع تعلق μ_{ij} از روی تابع هدف J که با توجه به محدودیت‌های موجود بر آن به‌صورت معادله لاگرانژ \bar{J} نوشته شد بیان می‌شود. این مراحل به ترتیب به‌صورت زیر است:

$$\frac{\partial \bar{J}}{\partial \mu_{ij}} = 0$$

$$\sum_{k=1}^M m \mu_{ij}^{m-1} w_{jk}^\beta d^2(x_{ik} - v_{jk}) - \lambda = 0$$

$$\mu_{ij} = \left(\frac{\lambda}{m \sum_{k=1}^M w_{jk}^\beta d^2(x_{ik} - v_{jk})} \right)^{\frac{1}{m-1}} \quad (11)$$

حال فقط بایستی λ را حساب کنیم. برای به دست آوردن λ به‌صورت زیر عمل می‌کنیم:

$$\frac{\partial \bar{J}}{\partial \mu_{il}} = 0$$

$$\sum_{k=1}^M m \mu_{il}^{m-1} w_{lk}^\beta d^2(x_{ik} - v_{lk}) - \lambda = 0$$

$$m \mu_{il}^{m-1} \sum_{k=1}^M w_{lk}^\beta d^2(x_{ik} - v_{lk}) = \lambda \quad (12)$$

از روی رابطه (*) و (** نتیجه می‌گیریم:

$$\mu_{ij} = \left(\frac{m \mu_{il}^{m-1} \sum_{k=1}^M w_{lk}^\beta d^2(x_{ik} - v_{lk})}{m \sum_{k=1}^M w_{jk}^\beta d^2(x_{ik} - v_{jk})} \right)^{\frac{1}{m-1}}$$

$$\sum_{l=1}^C \mu_{il} = \sum_{l=1}^C \mu_{ij} \left(\frac{\sum_{k=1}^M w_{lk}^\beta d^2(x_{ik} - v_{lk})}{\sum_{k=1}^M w_{jk}^\beta d^2(x_{ik} - v_{jk})} \right)^{\frac{1}{m-1}}$$

$$\mu_{ij} = \frac{1}{\sum_{l=1}^C \left(\frac{\sum_{k=1}^M w_{lk}^\beta d^2(x_{ik} - v_{lk})}{\sum_{k=1}^M w_{jk}^\beta d^2(x_{ik} - v_{jk})} \right)^{\frac{1}{m-1}}}$$

که در اینجا $d^2(x_{ik} - v_{jk})$ برابر است با $(x_{ik} - v_{jk})^2$. بدین ترتیب تابع تعلق μ_{ij} در فرمول (۶) به دست می‌آید.

در ادامه مراحل به دست آوردن فرمول به‌روزرسانی v_{jk} را خواهیم دید.

$$\frac{\partial \bar{J}}{\partial v_{jk}} = 0$$

$$\sum_{i=1}^N \mu_{ij}^m w_{jk}^\beta (+2)(\gamma_k)(x_{ik} - v_{jk}) \exp(-\gamma_k (x_{ik} - v_{jk})^2) = 0$$

$$\sum_{i=1}^N \mu_{ij}^m x_{ik} \exp(-\gamma_k (x_{ik} - v_{jk})^2) - \sum_{i=1}^N \mu_{ij}^m v_{jk} \exp(-\gamma_k (x_{ik} - v_{jk})^2) = 0$$

$$v_{jk} = \frac{\sum_{i=1}^N \mu_{ij}^m \exp(-\gamma_k (x_{ik} - v_{jk})^2) x_{ik}}{\sum_{i=1}^N \mu_{ij}^m \exp(-\gamma_k (x_{ik} - v_{jk})^2)}$$

بدین ترتیب فرمول به‌روزرسانی v_{jk} در معادله (۷) به دست می‌آید.

در مورد w_{jk} باید گفت اگر $D_{jk} = 0$ باشد و m_k تعداد صفاتی مانند l باشد که D_{jl} به ازای آن‌ها صفر است آنگاه w_{jk} به‌صورت $\frac{1}{m_k}$ تعریف می‌شود. از طرفی اگر $D_{jk} \neq 0$ اما صفتی مانند l وجود داشته باشد که D_{jl} به ازای آن صفر است و مراحل به دست آوردن فرمول به‌روزرسانی w_{jk} در حالتی که به ازای هر صفت از مجموعه داده، مجموع فواصل درون‌خوشه‌ای با توجه به این صفت مخالف صفر است به‌صورت زیر خواهد بود:

$$\frac{\partial \bar{J}}{\partial w_{jk}} = 0$$

$$\sum_{i=1}^N \mu_{ij}^m \beta w_{jk}^{\beta-1} d^2(x_{ik} - v_{jk}) - \alpha = 0$$

$$w_{jk} = \left(\frac{\alpha}{\beta \sum_{i=1}^N \mu_{ij}^m d^2(x_{ik} - v_{jk})} \right)^{\frac{1}{\beta-1}} \quad (13)$$

حال فقط بایستی α را حساب کنیم. برای به دست آوردن α به‌صورت زیر عمل می‌کنیم:

$$\frac{\partial \bar{J}}{\partial w_{jl}} = 0$$

$$\sum_{i=1}^N \mu_{ij}^m d^2(x_{ik} - v_{jk}) - \alpha = 0$$

$$\beta w_{jl}^{\beta-1} \sum_{i=1}^N \mu_{ij}^m d^2(x_{ik} - v_{jk}) = \alpha \quad (14)$$

از روی رابطه (#) و (##) نتیجه می‌گیریم:

[11] M. Yamamoto and Y. Terada, "Functional factorial - means analysis," *Computational Statistics & Data Analysis*, vol. 79, pp. 133-148, 2014.

[12] M.C. Naldi and R.J.G.B. Campello, "Evolutionary k-means for distributed data sets," *Neurocomputing*, vol. 127, pp. 30-42, 2014.

[13] Z. Ji, Y. Xia, Q. Sun, and G. Cao, "Interval-valued possibilistic fuzzy C-means clustering algorithm," *Fuzzy Sets and Systems*, vol. 253, pp. 138-156, 2014.

[14] L. Szilágyi and S.M. Szilágyi, "Generalization rules for the suppressed fuzzy c-means clustering algorithm," *Neurocomputing*, vol. 139, pp. 298-309, 2014.

[15] D. Zhang, M. Ji, J. Yang, Y. Zhang, and F. Xie, "A novel cluster validity index for fuzzy clustering based on bipartite modularity," *Fuzzy Sets Syst.*, vol. 253, pp. 122-137, 2014.

[16] M.A. Khalilia, J. Bezdek, M. Popescu, and J.M. Keller, "Improvements to the relational fuzzy c-means clustering algorithm," *Pattern Recognition*, vol. 47, pp. 3920-3930, 2014.

[17] E.Y. Chan, W.K. Ching, M.K. Ng, and J.Z. Huang, "An optimization algorithm for clustering using weighted dissimilarity measures," *Pattern Recognition*, vol. 37, pp. 943-952, 2004.

[18] H. Frigui and O. Nasraoui, "Unsupervised learning of prototypes and attribute weights," *Pattern Recognition*, vol. 37, pp. 567-581, 2004.

[19] W.S.T. GAO Jun, "Fuzzy Clustering Algorithm with Ranking Features and Identifying Noise Simultaneously," *Acta Automatica Sinica*, vol. 35, pp. 145-153, 2009.

[20] J.Z. Huang, M.K. Ng, R. Hongqiang, and L. Zichen, "Automated variable weighting in k-means type clustering," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, pp. 657-668, 2005.

[21] W.L. Hung, M.S. Yang, and D.H. Chen, "Bootstrapping approach to feature-weight selection in fuzzy c-means algorithms with an application in color image segmentation," *Pattern Recognition Letters*, vol. 29, pp. 1317-1325, 2008.

[22] J. Liping, M.K. Ng, and J.Z. Huang, "An Entropy Weighting k-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 19, pp. 1026-1041, 2007.

[23] X. Wang, Y. Wang, and L. Wang, "Improving fuzzy c-means clustering based on feature-weight learning," *Pattern Recognition Letters*, vol. 25, pp. 1123-1132, 7/16/ 2004.

[24] H.J. Xing and M.H. Ha, "Further improvements in Feature-Weighted Fuzzy C-Means," *Information Sciences*, vol. 267, pp. 1-15, 2014.

[25] X.B. Zhi, J.L. Fan, and F. Zhao, "Robust local feature weighting hard c-means clustering algorithm," *Neurocomputing*, vol. 134, pp. 20-29, 2014.

[26] K.L. Wu and M.S. Yang, "Alternative c-means clustering algorithms," *Pattern Recognition*, vol. 35, pp. 2267-2278, 2002.

[27] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Springer US, 1981.

$$w_{jk} = \frac{\beta w_{jl} \beta^{-1} \sum_{i=1}^N \mu_{ij}^m d^2(x_{ik} - v_{jk}) \frac{1}{\beta^{-1}}}{\beta \sum_{i=1}^N \mu_{ij}^m d^2(x_{ik} - v_{jk})}$$

$$w_{jk} = \frac{w_{jl}}{\sum_{i=1}^N \mu_{ij}^m d^2(x_{ik} - v_{jk}) \frac{1}{\beta^{-1}}}$$

$$\sum_{l=1}^C w_{jl} = \sum_{l=1}^M w_{jk} \left(\frac{\sum_{i=1}^N \mu_{ij}^m d^2(x_{ik} - v_{jk}) \frac{1}{\beta^{-1}}}{\sum_{i=1}^N \mu_{ij}^m d^2(x_{ik} - v_{jk})} \right)^{\beta^{-1}}$$

$$w_{jk} = \frac{1}{\sum_{l=1}^M \left(\frac{\sum_{i=1}^N \mu_{ij}^m d^2(x_{ik} - v_{jk}) \frac{1}{\beta^{-1}}}{\sum_{i=1}^N \mu_{ij}^m d^2(x_{ik} - v_{jk})} \right)^{\beta^{-1}}}$$

بدین ترتیب فرمول به‌روزرسانی w_{jk} در معادله (۸) به دست می‌آید.

مراجع

[1] D.G. Calò, A. Montanari, and C. Viroli, "A hierarchical modeling approach for clustering probability density functions," *Computational Statistics & Data Analysis*, vol. 71, pp. 79-91, 2014.

[2] C.Z. Li, Z.B. Xu, and T. Luo, "A heuristic hierarchical clustering based on multiple similarity measurements," *Pattern Recognition Letters*, vol. 34, pp. 155-162, 2013.

[3] K. Liao, G. Liu, L. Xiao, and C. Liu, "A sample-based hierarchical adaptive K-means clustering method for large-scale video retrieval," *Knowledge-Based Systems*, vol. 49, pp. 123-133, 2013.

[4] S. M. Razavi Zadegan, M. Mirzaie, and F. Sadoughi, "Ranked k-medoids: A fast and accurate rank-based partitioning algorithm for clustering large datasets," *Knowledge-Based Systems*, vol. 39, pp. 133-143, 2013.

[5] H. Jiang, J. Li, S. Yi, X. Wang, and X. Hu, "A new hybrid method based on partitioning-based DBSCAN and ant clustering," *Expert Systems with Applications*, vol. 38, pp. 9373-9381, 2011.

[6] Y.J. Wang, "A clustering system for data sequence partitioning," *Expert Syst. Appl.*, vol. 38, pp. 659-666, 2011.

[7] A. Hatamlou, "In search of optimal centroids on data clustering using a binary search algorithm," *Pattern Recognition Letters*, vol. 33, pp. 1756-1760, 2012.

[8] M.R.P. Ferreira and F.D.A.T. de Carvalho, "Kernel-based hard clustering methods in the feature space with automatic variable weighting," *Pattern Recognition*, vol. 47, pp. 3082-3095, 2014.

[9] F.D.A.T. de Carvalho, Y. Lechevallier, and F.M. deMelo, "Partitioning hard clustering algorithms based on multiple dissimilarity matrices," *Pattern Recognition*, vol. 45, pp. 447-464, 2012.

[10] J.M. Górriz, J. Ramírez, E.W. Lang, and C.G. Puntonet, "Hard C-means clustering for voice activity detection," *Speech Communication*, vol. 48, pp. 1638-1649, 2006.

زیرنویس‌ها

- ۱Hierarchical
- ۲Partitioning
- ۳Dendrogram
- ۴Hard
- ۵Exclusive
- ۶Soft
- ۷Fuzzy
- ۸Hard C-Means
- ۹Fuzzy C-Means
- ۱۰Xiao
- ۱۱Outlier
- ۱۲<http://archive.ics.uci.edu/ml>
- ۱۳Robust Local Feature Weighted Fuzzy C-Mean
- ۱۴Karush-Kuhn-Tucker