چشم‌انداز مدیریت صنعتی
سال نهم، شماره ۳۶، زمستان ۱۳۹۸
شاپا چاپی: ۹۸۷۴-۲۲۵۱، شاپا الکترونیکی: ۴۱۶۵-۲۶۴۵
ص‌ص ۶۱ – ۴۱

توسعه یک روش هوشمند خوشه‌بندی چندمعیاره مبتنی بر پرامتی

امیر دانشور*، مهدی همایون‌فر، آنیا فرهمندنژاد*****

چکیده

در سال‌های اخیر مسئله جدیدی با عنوان«خوشه‌بندی چندمعیاره» ظهور کرده که هدف آن، دسته‌بندی گزینه‌ها در گروه‌های همگنی به نام خوشه با توجه به معیارهای ارزیابی متفاوت است. در ادامه پژوهش‌های انجام‌گرفته در مبانی نظری، پژوهش حاضر با ترکیب الگوریتم K– میانگین و تکنیک پرامتی، به‌دنبال توسعه یک روش جدید خوشه‌بندی چندمعیاره است. پارامترهای مسئله، پروفایل‌های جداکننده خوشه‌ها هستند که برای بهینه‌سازی آن‌ها از الگوریتم ژنتیک استفاده شده است. برای تنظیم پارامترهای ژنتیک نیز از روش تاگوچی استفاده می‌شود. در این مدل‌سازی، متغیرها در هر مرحله از به‌روزرسانی جواب‌ها، با توجه به فاصله امتیاز جریان خالص خود از پروفایل‌ها به نزدیک‌ترین خوشه تخصیص می‌یابند. عملگر جهش نیز صرفاً زمانی اعمال می‌شود که میزان شباهت کروموزوم‌ها در هر جمعیت به حد خاصی برسد که این هوشمندسازی موجب کاهش زمان محاسباتی شده است. درنهایت با اجرای روش پیشنهادی بر روی چند نمونه مسائل تصادفی مالی، عملکرد آن با سایر الگوریتم‌های شناخته‌شده خوشه‌بندی مقایسه شده است. نتایج نشان می‌دهد که روش پیشنهادی ضمن تعیین تعداد بهینه خوشه‌ها، در مقایسه با سایر الگوریتم‌ها، جواب‌های دقیق‌تری ارائه می‌دهد.

کلیدواژه‌ها: خوشه‌بندی چندمعیاره؛ الگوریتم ژنتیک؛ الگوریتم K–میانگین؛ شاخص سیلوئت؛ پرامتی.

تاریخ دریافت مقاله: ۱۳۹۷/۰۹/۲۴، تاریخ پذیرش مقاله: ۱۳۹۸/۱۰/۰۸.

* استادیار مدیریت صنعتی، واحد الکترونیکی، دانشگاه آزاد اسلامی، تهران، ایران.
** استادیار مدیریت صنعتی، واحد رشت، دانشگاه آزاد اسلامی، رشت، ایران (نویسنده مسئول).
E-mail: homayounfar@iaurasht.ac.ir
*** کارشناسی ارشد مدیریت صنعتی، واحد الکترونیکی، دانشگاه آزاد اسلامی، تهران، ایران.

۱. مقدمه

در دهه‌های اخیر، تحقیق در عملیات کارایی بالای خود را در حل تعداد زیادی از مسائل، اثبات کرده است. در این میان تصمیم‌های استراتژیکی که بر پایه معیارهای متعدد و متناقض اتخاذ می‌شوند از اهمیت و توجه ویژه‌ای برخوردارند. در مواجهه با این مسائل الگوریتم‌هایی از قبیل: تئوری مطلوبیت چند شاخصه[1] فرایند تحلیل سلسله‌مراتبی[2] الکتر، مکبت[3] و پرامتی[4] برای ساختاردهی به فرآیند تصمیم‌گیری ارائه شده‌اند [۱۶، ۴۰، ۱۹، ۱۲، ۸]. مسائل تصمیم‌گیری را می‌توان به‌صورت مجموعه‌ای از متغیرها که بر اساس چندین معیار متعارض ارزیابی می‌شوند، مدل‌سازی کرد. پژوهشگران معمولاً سه رویکرد اساسی در برخورد با این مسئله اتخاذ می‌کنند: در رویکرد نخست، انتخاب بهترین گزینه‌ها موردنظر است؛ رویکرد دوم به رتبه‌بندی گزینه‌ها از بهترین‌ها تا بدترین اشاره دارد و در نهایت، رویکرد سوم به طبقه‌بندی گزینه‌ها در دسته‌های از پیش تعریف‌شده می‌پردازد [۹]. در سال‌های اخیر، دسته‌ای جدید از مدل‌ها با عنوان «مدل‌های خوشه‌بندی چندمعیاره» در حوزه تصمیم‌گیری موردتوجه قرار گرفتند که هدف آن‌ها، دسته‌بندی مجدد گزینه‌ها در گروه‌های همگن (خوشه‌ها)، هم‌زمان با بهینه‌سازی معیارهای مختلف است [۴۱].

عناصر قرارگرفته در هر خوشه بیشترین شباهت را با یکدیگر و کمترین تشابه را با عناصر دیگر خوشه‌ها دارند. ازطرف‌دیگر در حوزه تحلیل داده، دو رویکرد اساسی برای گروه‌بندی داده‌ها وجود دارد که عبارت‌اند از: طبقه‌بندی و خوشه‌بندی. طبقه‌بندی یک گروه‌بندی نظارت‌شده است که داده‌ها را در گروه‌های ازپیش‌تعریف‌شده طبقه‌بندی می‌کند که به «کلاس» معروف هستند. خوشه‌بندی یک گروه‌بندی بدون نظارت است که دانشی درباره ساختار داده موجود نیست و هدف کشف ساختار داده‌ها بااستفاده‌از سنجه‌هایی مانند شباهت است؛ بنابراین در داخل خوشه‌ها عناصر مشابه یکدیگر و مابین خوشه‌ها عناصر متفاوت از یکدیگر هستند [۲۲]. شباهت‌های خاصی بین طبقه‌بندی و خوشه‌بندی از حوزه تحلیل داده‌ها و دسته‌بندی و رتبه‌بندی از حوزه تصمیم‌گیری چندشاخصه[5](MADM) وجود دارد؛ اما به مسئله خوشه‌بندی در حوزه MADM توجه چندان زیادی نشده است. در حوزه تحلیل داده‌ها چندین روش برای خوشه‌بندی توسعه داده شده است که صرفاً از سنجه‌های فاصله برای اندازه‌گیری میزان تشابه دو متغیر استفاده می‌شود؛ ولی هیچ‌یک از این روش‌ها از اطلاعات فراوانی که MADM در اختیار قرار می‌دهد، استفاده نمی‌کنند. این اطلاعات شامل ترجیحات تصمیم‌گیرندگان درباره متغیرها است [۳۴].

1. Multi Attribute Utility Theory (MAUT)
2. Analytical Hierarchy Process (AHP)
3.MACBETH
4. PROMETHEE
5. Multi Attribute Decision Making

در این پژوهش، مسئله خوشه‌بندی داده‌ها در حوزه مبحث MADM تعریف و یک روش برای حل آن ارائه خواهد شده است. مقاله در ادامه بدین شکل ساختاربندی می‌شود که در بخش دوم، روش‌های شناخته‌شده خوشه‌بندی، هم در حوزه تحلیل داده‌ها و هم حوزه MADM به‌طورکلی مرور می‌شود؛ سپس در بخش سوم، مسئله خوشه‌بندی در MADM مبتنی بر روابط باینری بین متغیرها، تعریف و یک الگوریتم برای حل این مسئله پیشنهاد می‌شود. در بخش چهارم، این رویکرد با آزمایش روی مسائل نمونه و همچنین مقایسه با روش‌های موجود در مبانی نظری، اعتبارسنجی می‌شود و درنهایت در بخش پنجم نتایج پژوهش موردبررسی قرار می‌گیرد.

۲. مبانی نظری و پیشینه پژوهش

در مبانی نظری خوشه‌بندی، مدل‌های مختلفی ارائه شده‌اند که در زمینه‌های بسیاری از جمله شناسایی الگو، یادگیری ماشین، داده‌کاوی، بازیابی اطلاعات و انفورماتیک زیستی کاربرد داشته‌اند [۴۵]. کرینک و همکاران (۲۰۰۷) از ارزیابی تفاضلی برای خوشه‌بندی اعتباری استفاده کرده و عملکرد روش ارائه‌شده را با الگوریتم ژنتیک، ازدحام ذرات و جست‌وجوی تصادفی مقایسه کردند [۳۰]. اولسون و زوبی (۲۰۰۸) با به‌کارگیری مدل‌های خوشه‌بندی لاجیت، شبکه عصبی و نزدیک‌ترین همسایه K- میانگین و با تکیه بر ۲۶ نسبت مالی، به بررسی تفاوت‌های دو دسته از بانک‌ها پرداختند [۳۷]. لیائو و همکاران (۲۰۰۸) از یک رویکرد داده‌کاوی دومرحله‌ای متشکل از الگوریتم‌های آپریوری[1] و K- میانگین برای خوشه‌بندی سهام در بازار بورس تایوان استفاده کردند [۳۳]. روچا و همکاران (۲۰۱۳)، روشی برای خوشه‌بندی چندمتغیره پیشنهاد دادند که در آن از یک رویکرد خوشه‌بندی (ترجیحات تصمیم‌گیرنده را جمع‌بندی نمی‌کند) و یک روش چندمعیاره برای تعیین روابط بین گروهی استفاده شده است [۳۹].

یو و همکاران (۲۰۱۷)، یک الگوریتم سه‌سطحی K- میانگین و یک الگوریتم دوسطحی دولایه‌ای برای خوشه‌بندی داده‌ها ارائه کرده و کارایی آن را با روش K- میانگین استاندارد مقایسه کردند [۴۶]. ژای و همکاران (۲۰۱۷)، ویژگی‌های کلیدی گروه‌های همکاری[2] را با شبکه‌های مستقیم و غیرمستقیم در بازار مالی موردبررسی و تحلیل قرار دادند و در ادامه، معامله‌کنندگان و مبادلات را با توجه به الگوریتم K- میانگین خوشه‌بندی کردند [۴۸]. کاپو و همکاران (۲۰۱۷)، تقریبی کارا از خوشه‌بندی K- میانگین برای داده‌های با حجم بالا ارائه دادند [۱۰]. اسلام و همکاران (۲۰۱۸) با ترکیب روش K- میانگین و الگوریتم ژنتیک و اصلاح کروموزوم‌های ژنتیک، یک تکنیک خوشه‌بندی جدید ارائه کردند [۲۵].

1. Apriori
2. Collusive Clique

یقینی و ورد (۲۰۱۲) با بهره‌گیری از روش اندازه‌گیری فاصله میان مقادیر دسته‌ای، روشی جدید مبتنی بر الگوریتم ژنتیک را برای خوشه‌بندی داده‌های مختلط ارائه دادند [۴۴]. حامدی و همکاران (۲۰۱۳) به‌منظور شناخت انواع مشتریان یک شرکت غذایی و خوشه‌بندی آن‌ها از دو معیار ارزش‌گذاری LRFM[1] و RFM[2] استفاده کردند [۲۱]. علی‌حیدری بیوکی و خادمی زارع (۲۰۱۵)، پس از شناسایی معیارهای اعتباردهی به متقاضیان حقوقی دریافت تسهیلات بانکی، با استفاده از تکنیک بهبودیافته تحلیل پوششی داده‌ها، روشی کارآمد برای دسته‌بندی مشتریان حقوقی ارائه دادند [۲]. قربانپور و همکاران (۲۰۱۵) از تلفیق الگوریتم‌های ژنتیک و C- میانگین در محیط فازی برای خوشه‌بندی مشتریان «بانک رفاه» شهر تهران و غلبه بر مشکلاتی مانند حساس‌بودن به مقدار اولیه و گرفتارشدن در دام بهینه محلی استفاده کردند [۲۰].

لای و همکاران (۲۰۱۸) از یک رویکرد ترکیبی متشکل از الگوریتم خوشه‌بندی داده‌ها، درخت تصمیم فازی و الگوریتم ژنتیک برای ایجاد یک سیستم تصمیم مبتنی بر داده‌های تاریخی استفاده کردند [۳۲]. ژائو و همکاران (۲۰۱۸) یک نسخه اصلاح‌شده K- میانگین (VRKM[3]++) که از کارایی بیشتری نسبت به مدل‌های K- میانگین و VRKM برخوردار است، برای خوشه‌بندی داده‌ها ارائه دادند [۵۰]. کومار (۲۰۱۸) یک رویکرد ترکیبی متشکل از الگوریتم K- میانگین و الگوریتم شیر مورچه را برای خوشه‌بندی مجموعه داده‌های مالی به‌کار برد [۳۱]. ژانگ و همکاران (۲۰۱۸) به‌منظور تجزیه‌وتحلیل داده‌های حاصل از منابع و نگرش‌های چندگانه و خوشه‌بندی آن‌ها، یک رویکرد دومرحله‌ای موزون مشارکتی بر مبنای الگوریتم K- میانگین ارائه دادند [۴۹].

خدیور و حامدی (۲۰۱۵) از الگوریتم‌های خوشه‌بندی و قواعد انجمنی برای تدوین الگوی سیاست‌گذاری تخفیف‌دهی مناسب به مشتریان در «شرکت پخش مواد غذایی پگاه» استفاده کردند [۲۸]. زارع احمدآبادی و همکاران (۲۰۱۶) با استفاده از الگوریتم بهینه‌سازی مورچگان به خوشه‌بندی بازار یک شرکت تولیدکننده کاشی در ایران پرداختند [۴۷]. فائزی‌راد و پویا (۲۰۱۶) با استفاده از تحلیل داده‌های ۳۰۰۰ فروشگاه آنلاین و بر اساس شاخص‌های موردنظر تأمین‌کننده، از شبکه عصبی مصنوعی SOM[4] و الگوریتم K- میانگین به‌منظور خوشه‌بندی فروشگاه‌های آنلاین استفاده کردند [۱۷]. عزیزی و بلاغی اینانلو (۲۰۱۶)، انتظارات کاربرانِ همراه‌بانک کشاورزی را در قالب ۴ عامل اصلی شامل خدمات اصلی، خدمات افزوده، شرایط تسهیل‌کننده کاربری و ادراک مفیدبودن، بررسی کردند. در مرحلهٔ بعد، پس از مشخص‌کردن تعداد بهینهٔ

1. Including Length, Recency, Frequency, and Monetary indices
2. Including Recency, Frequency, and Monetary indices
3. Variance Reduction K-Means
4. Self-Organizing Map

خوشه‌ها با روش وارد[1]از مدل K- میانگین برای خوشه‌بندی داده‌ها استفاده شد [۶]. ایمانی و عباسی (۲۰۱۷) با بررسی تراکنش‌های ثبت‌شده از مشتریان «فروشگاه رفاه» شهر زاهدان، پس از تعیین مقادیر RFM، تعداد بهینه خوشه‌ها را محاسبه کردند. در مرحله بعد مشتریان با الگوریتم فازی C-میانگین به هفت خوشه تقسیم شدند [۲۴]. علیزاده و پویا (۲۰۱۷) در مطالعه‌ای با استفاده از روش خوشه‌بندی سلسله‌مراتبی به ارزیابی و خوشه‌بندی ۳۱ بانک و مؤسسه مالی و اعتباری ایران بر اساس ۶ شاخص ترافیکی وب‌سایت پرداختند [۳]. نبی‌لو و دانش‌پور (۲۰۱۷) با استفاده از ترکیب معیارهای شباهت اورلای و جاکارد[2]بر روی یک الگوریتم سلسله‌مراتبی، رویکردی را برای خوشه‌بندی داده‌های دسته‌ای پیشنهاد دادند [۳۶]. کشاورز حدادها و همکاران (۲۰۱۸) با استفاده از الگوریتم K- میانگین، مدلی برای خوشه‌بندی، ارزیابی و انتخاب پروژه‌ها ارائه کردند [۲۷].

احمدزاده گلی و همکاران (۲۰۱۸) به ارائه خوشه‌بندی K- میانگین لاینکس هوشمند که مدل توسعه‌یافته خوشه‌بندی K- میانگین است، پرداختند [۱]. خدیور و مجیبیان (۲۰۱۸) با تلفیق فرآیند داده‌کاوی و روش‌های تصمیم‌گیری چندشاخصه، روشی ترکیبی (فرایند تحلیل سلسله مراتبی، K- میانگین و شبکه عصبی) را برای خوشه‌بندی کارگاه‌های صنعتی ارائه دادند [۲۹]. با مطالعه پژوهش‌های پیشین، مشاهده می‌شود که الگوریتم ارائه‌شده از نظر تئوریک، مبنای جدیدی برای خوشه‌بندی ارائه می‌دهد.

دی‌اسمت و گوژمان (۲۰۱۲) الگوریتم K- میانگین را در فضای MADM توسعه دادند [۱۴]. این نخستین تلاش برای خوشه‌بندی در MADM بود که از عباراتی همچون بی‌تفاوتی،[3] ارجحیت قوی[4]و غیرقابل‌مقایسه‌بودن[5]که ذاتاً متعلق به این حوزه از علم است، استفاده می‌کرد. همچنین مطالعه آنها یک نسخه توسعه یافته از رویکرد شناخته‌شده و پرکاربرد K- میانگین بود. این مسئله بعداً در دی اسمت (۲۰۱۳) با عنوان «روش P2CLUST» موردمطالعه قرار گرفت [۱۳]. در هر دو مطالعه، پژوهشگران روابط برتری بین متغیرها را به‌صورت قطعی در نظر گرفته بودند. فیگورا و همکاران (۲۰۰۵) نیز الگوریتم K- میانگین را در یک چارچوب چندمعیاره توسعه دادند [۱۹]. جدیدترین پژوهش درباره توسعه این الگوریتم کلاسیک را می‌توان در بارودی و سافیا (۲۰۱۰) یافت [۷]. دی اسمت و همکاران (۲۰۱۲) یک رویکرد خوشه‌بندی پیشنهاد کردند که یک مجموعه خوشه رتبه‌بندی شده پیدا می‌کرد [۱۵].

1. Ward
2. Overlay and Jaccard
3. Indifference
4. Strict Preference
5. Incomparability

سارازین و همکاران (۲۰۱۸)، یک مدل خوشه‌بندی مبتنی بر روش پرامتی I ارائه دادند و به مقایسه عملکرد مدل پیشنهادی با مدل P2CLUST پرداختند [۴۱]. جدیدترین پژوهش در این موضوع به‌وسیله فرناندز و همکاران (۲۰۱۰) انجام شده است [۱۸]. در این رویکردها، رتبه بین خوشه‌ها کامل است؛ اما روچا و همکاران (۲۰۱۳) اخیراً روی روشی کار کرده‌اند که مجموعه خوشه‌هایی را می‌یابد که به‌صورت بخشی، رتبه‌بندی می‌شوند [۳۹].

مرور پژوهش‌های یادشده نشان می‌دهد که در ترکیب K- میانگین و پرامتی، نقایص الگوریتم اصلی مانند لزوم تعیین تعداد خوشه‌ها قبل از خوشه‌بندی، حساسیت به مرحله شروع اولیه که ممکن است باعث قرارگرفتن در بهینه محلی شود و همچنین برخی محدودیت‌های مربوط به سنجه محاسبه فاصله همچنان پابرجا است. در فضای چندمعیاره، تعبیری که از فاصله بین دو متغیر می‌شود واقعاً متناسب نیست؛ زیرا کلیه معیارها باید بهینه‌سازی شوند. بنابراین مقایسه بین دو متغیر اغلب دشوار بوده و بهتر است روابط ارجحیت باینری[1] بین متغیرها در نظر گرفته شود [۴۱]؛ ازاین‌رو روش چندمعیاره پرامتی که این روابط را برای تحلیل مسئله مورداستفاده قرار می‌دهد، در این پژوهش به‌کار می‌رود. این روش به‌خاطر سادگی استفاده، وجود نرم‌افزارهای کاربرپسند و گستره وسیعی از کاربردهای دنیای واقعی شناخته شده است [۹]. با توجه به موارد بیان‌شده، در این پژوهش سعی می‌شود تا با ارائه یک روش ترکیبی، نقاط ضعف روش‌های پیشین برطرف گردد و با به‌کارگیری الگوریتم ژنتیک، یک مدل بهینه خوشه‌بندی ارائه شود.

۳. روش‌شناسی پژوهش

در این پژوهش از تکنیک پرامتی و الگوریتم ژنتیک برای مدل‌سازی مسئله خوشه‌بندی استفاده شده است؛ بنابراین، پژوهش حاضر از نظر روش تحلیلی ـ ریاضی و از نظر هدف کاربردی است [۴۷]. به‌منظور ارزیابی کارایی الگوریتم پیشنهادی، مسائل نمونه تصادفی تولید می‌شوند که با توجه به تعداد خوشه‌ها (از ۲ تا ۸ خوشه) در ۷ گروه قرار می‌گیرند. معیارهای هر مسئله شامل، ۱۸ نسبت مالی (نسبت بدهی، نسبت دارایی‌های جاری و غیره) و متغیرها، ۴۶۰ شرکت فرضی هستند. با توجه به وزن شاخص‌ها، در هر دسته ۴ مسئله تصادفی تولید می‌شود؛ بنابراین درمجموع ۲۸ مسئله نمونه تصادفی شکل می‌گیرد؛ سپس هر یک از مسائل نمونه، ۵۰ بار تکرار شده و میانگین جواب‌های به‌دست‌آمده به‌عنوان نتیجه الگوریتم در نظر گرفته می‌شوند.

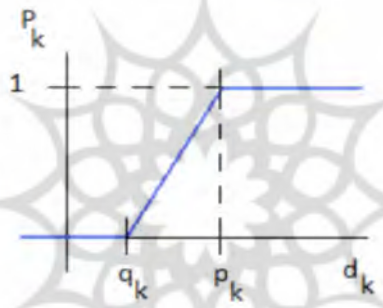
تکنیک بررسی و تحلیل داده‌ها. پرامتی یکی از تکنیک‌های شناخته‌شده در تصمیم‌گیری چندمعیاره است که نسخه‌های متنوعی از آن (پرامتی I، II، III، IV، V و VI) برای رتبه‌بندی

1. Binary Preference Relations

گزینه‌ها تحت شرایط مختلف و نیز طبقه‌بندی گزینه‌ها (پرامتی TRI و FlowSort) توسعه یافته است [۱۳]. نسخه‌های مختلف این روش در زمینه‌های متنوعی مانند انتخاب پورتفولیو [۴]، برون‌سپاری خدمات [۵]، تعمیرات و نگهداری [۱۱]، ارزیابی عملکرد زنجیره تأمین [۲۶]، انتخاب اعضای گروه پروژه [۳۸]، مدیریت آب [۴۳] و غیره به‌کار رفته است. یکی از پرکاربردترین نسخه‌های پرامیتی، تکنیک پرامتی II است که مبنای رتبه‌بندی گزینه‌ها در آن امتیاز ϕ است. فرض کنید $A=\{a_1,a_2,...,a_n\}$، مجموعه‌ای از n گزینه موجود و $F=\{f_1,f_2,...,f_q\}$ مجموعه‌ای از q معیار ارزیابی باشد و هدف مدل‌سازی، حداکثرسازی این معیارها است؛ بدین منظور ابتدا تفاوت میان داده‌ها بر اساس تمام معیارهای موردبررسی محاسبه می‌شود:

$$d_k(a_i,a_j)=f_k(a_i)-f_k(a_j)$$ رابطه (۱)

این امر موجب انجام مقایسات زوجی گزینه‌ها با توجه به هر یک از معیارها خواهد شد که باید برای انجام مقایسات بعدی توسط تابعی مناسب بی‌مقیاس شوند. به این منظور استفاده از تابع خطی $P_k : R \to [0,1]$ پیشنهاد می‌شود:



شکل ۱. مثالی از تابع ترجیح خطی

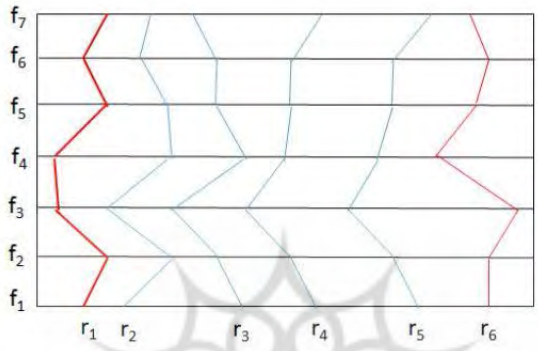
در ادامه با استفاده از رابطه ۲، با درنظرگرفتن وزن معیارها به هر مقایسه زوجی صورت‌پذیرفته یک مقدار (π) نسبت داده می‌شود:

$$\pi(a_i,a_j)=\sum_{k=1}^{q} w_k.P_k$$ رابطه (۲)

سپس امتیاز هر یک از گزینه‌های موجود با استفاده از رابطه ۳، محاسبه می‌شود:

$$\phi_A(a_i)=\frac{1}{n-1}\left\{\sum_{a_j\in A}[\pi(a_i,a_j)-\pi(a_j,a_i)]\right\}$$ رابطه (۳)

تا این مرحله از الگوریتم، گزینه‌ها با استفاده از تکنیک پرامتی II و با تخصیص مقادیر ϕ رتبه‌بندی شده‌اند. در ادامه به‌منظور تخصیص گزینه‌ها به دسته‌های همگن از الگوریتم P2CLUST استفاده می‌شود. در این الگوریتم فرض بر این است که K تعداد دسته‌های موردنظر برای تخصیص گزینه‌ها باشد ($C_1, C_2, \dots, C_k$)؛ همچنین دسته‌ی C_{i+1} نسبت به دسته‌ی C_i، ارجحیت دارد (این فرض عمومیت مسئله را تحت تأثیر قرار نخواهد داد). به هر یک از این دسته‌ها یک مرکز دسته (به‌صورت تصادفی) نسبت داده می‌شود که بر اساس فرض صورت‌پذیرفته، مرکز دسته مربوط به C_{j+1} نسبت به مرکز دسته مربوط به C_j در تمامی معیارها ارجحیت دارد (شکل ۲).



شکل ۲. وضعیت مرکز دسته‌ها نسبت به یکدیگر – شش دسته و هفت معیار

در این مرحله با روش تشریح‌شده برای تعیین مقدار ϕ برای گزینه‌ها، مقدار ϕ مرکز دسته‌های انتخاب‌شده محاسبه می‌شوند؛ سپس بر اساس بزرگی اختلاف میان ϕ هر یک از گزینه‌ها با ϕ مراکز دسته‌ها، نسبت به توزیع آن‌ها در دسته‌های موجود اقدام می‌شود. پس از پایان این مرحله و تخصیص تمام گزینه‌ها به دسته‌های موجود، مرکز دسته‌ها به‌روزرسانی شده و فرآیند تخصیص مجدداً اجرا خواهد شد و این روند مادامی‌که نحوه تخصیص گزینه‌ها دست‌خوش تغییر نشود، ادامه می‌یابد.

همان‌گونه که تشریح شد، مقادیر ϕ تأثیر بسزایی در تعیین سطح کیفیت خوشه‌بندی دارد؛ بنابراین در ادامه یک روش ترکیبی پیشنهاد می‌شود که قادر است در یک فرآیند هوشمند، بهترین مقادیر ϕ را برای دستیابی به باکیفیت‌ترین خوشه‌بندی بیابد. در روش پیشنهادی ابتدا مقادیر تصادفی برای ϕ تولید شده و سپس بر اساس این مقادیر الگوریتم خوشه‌بندی اجرا شده و شاخص سیلوئت به‌عنوان میزان کارایی خوشه‌بندی تعیین می‌شود. با توجه به شاخص سیلوئت، مجدداً مقادیر جدیدی برای ϕ تولید شده و داده‌ها خوشه‌بندی می‌شوند. این فرآیند تا جایی ادامه

می‌یابد که به خوشه‌بندی با بالاترین کارایی دست یافته شود. گام‌های الگوریتم پیشنهادی به‌صورت زیر است:

گام ۱: تولید مقادیر تصادفی ϕ برای تمامی گزینه‌ها؛
گام ۲: انتخاب مرکز دسته‌ها به‌طور تصادفی؛
گام ۳: تکرار گام‌های چهارم تا هفتم به تعداد معین؛
گام ۴: تخصیص گزینه‌ها به خوشه‌ها بر اساس مقادیر ϕ؛
گام ۵: به‌روزرسانی مرکز دسته‌ها؛
گام ۶: محاسبه شاخص سیلوئت؛
گام ۷: محاسبه مقادیر ϕ برای مرکز خوشه‌ها.

–ارزیابی کروموزوم‌ها (شاخص سیلوئت). کیفیت خوشه‌ها در روش‌های یادگیری بدون نظارت از طریق روش‌های ارزیابی درونی انجام می‌شود که ارزیابی می‌کنند خوشه‌ها تا چه حد از هم جدا و تا چه حد به هم فشرده هستند. یک نمونه از این شاخص‌ها شاخص سیلوئت است. فرض کنید داده‌ها توسط روشی دلخواه همانند K– میانگین در K خوشه، دسته‌بندی شده‌اند. برای هر داده i، $a(i)$ به‌عنوان میانگین عدم‌تشابه داده i با سایر اعضای متعلق به دسته این داده در نظر گرفته می‌شود.

بدیهی است که هر چه این مقدار کوچک‌تر باشد، تخصیص صورت‌پذیرفته مطلوب‌تر خواهد بود؛ همچنین $b(i)$ به‌عنوان کمترین میانگین عدم‌تشابه داده i با خوشه‌های دیگر (که به آن‌ها تعلق ندارد) در نظر گرفته می‌شود. خوشه‌ای که $b(i)$ به آن تعلق دارد «خوشه همسایه» نامیده می‌شود؛ چراکه بهترین خوشه بعدی برای تخصیص داده‌ی i است. با توجه به توضیحات بیان‌شده، شاخص سیلوئت به‌صورت زیر محاسبه می‌شود.

$$s(i)=\frac{b(i)-a(i)}{\max\{a_i,b_i\}}$$ رابطه (۴)

رابطه ۵، شکل گسترده رابطه ۴ را نشان می‌دهد.

$$s(i)=\begin{cases}1-a(i)/b(i) & if \quad a(i)<b(i)\\ 0 & if \quad a(i)=b(i)\\ b(i)/a(i)-1 & if \quad a(i)>b(i)\end{cases}$$ رابطه (۵)

بر اساس رابطه ۶ می‌توان گفت که اندازه شاخص $s(i)$ همواره عددی بین ۱- تا ۱+ است. برای $s(i)$ نزدیک به ۱، $a(i)<<b(i)$ موردنیاز خواهد بود. ازآنجاکه $a(i)$، معیاری برای اندازه‌گیری عدم‌تعلق داده i به خوشه تخصیص‌یافته خود است (مقدار کوچک آن حاکی از تخصیص مناسب داده به خوشه است) و مقدار بزرگ $b(i)$ حکایت از عدم‌تعلق داده i به خوشه همسایه خود دارد، هر چه مقدار $s(i)$ به ۱ نزدیک‌تر باشد، کرموزوم از میزان شایستگی بالاتری برخوردار بوده و تخصیص مناسب‌تری صورت گرفته است. با منطق مشابه می‌توان نتیجه گرفت مقدار نزدیک به ۱- برای شاخص سیلوئت نشان‌دهنده آن است که تخصیص داده موردبررسی به خوشه همسایه انتخاب بهتری خواهد بود. درنهایت مقدار نزدیک به صفر به‌معنای قرارگرفتن داده موردنظر در مرز دو خوشه است و تغییر خوشه تأثیر چندانی در کیفیت خوشه‌بندی نخواهد داشت.

- تطبیق الگوریتم ژنتیک با مسئله خوشه‌بندی چندمعیاره مبتنی بر تکنیک پرامتی. به‌منظور تعیین مقادیر ϕ، در این پژوهش از الگوریتم ژنتیک که در سال ۱۹۷۵ توسط هلند[1][۲۳] توسعه یافته است، استفاده می‌شود. مبنای الگوریتم ژنتیک، استفاده از جست‌وجوی تصادفی برای بهینه‌سازی مسائل و فرآیندهای یادگیری است. الگوریتم ژنتیک می‌تواند نواحی مختلف فضای جواب را هم‌زمان جست‌وجو کند. مشکل اصلی این الگوریتم، همگرایی به بهینه محلی است که به‌دلیل مشابهت کروموزوم‌ها رخ می‌دهد. راهکار اصلی الگوریتم برای فرار از بهینه محلی استفاده از عملگر جهش است. به‌منظور بهبود کارایی الگوریتم بهتر است تا عملگر جهش زمانی اعمال شود که شباهت کروموزوم‌ها بسیار زیاد است. این پژوهش با ارائه الگوریتم ژنتیک با جهش ابتکاری[2]که عملگر جهش در آن تحت شرایطی خاص اعمال می‌شود، موجب بهبود عملکرد الگوریتم ژنتیک می‌شود.

کروموزوم: در الگوریتم ژنتیک هر کروموزوم نشان‌دهنده یک نقطه در فضای جست‌وجو و یک راه‌حل ممکن برای مسئله موردنظر است. کروموزوم‌ها از تعداد ثابتی ژن (متغیر) تشکیل می‌شوند و برای نمایش آن‌ها معمولاً از کدگذاری استفاده می‌شود. در این پژوهش کروموزوم باید مقادیر ϕ را به‌عنوان پارامتر ورودی روش P2CLUST، نشان دهد؛ بدین‌منظور از یک کروموزوم کدگذاری شده استفاده شده است. این کروموزوم یک ماتریس M*N است که در آن M، تعداد خوشه‌ها و N، تعداد شاخص‌ها است. تمامی اعداد موجود در این ماتریس، تصادفی و بین صفر و یک هستند و اعداد موجود در هر سطر مربوط به یک خوشه است.

1. Holland
2. Heuristic Genetic Algorithm (HGA)

عملکرد تقاطع: در جریان عمل تقاطع، بخش‌هایی از کروموزوم‌ها به‌صورت اتفاقی جایگزین یکدیگر می‌شوند که این مسئله باعث می‌شود فرزندان ترکیبی از خصوصیات والدین خود را به‌همراه داشته باشند. در پژوهش حاضر نحوه کار عملگر تقاطع به این صورت است که دو عضو از مجموعه جواب به‌عنوان کروموزوم‌های والد انتخاب می‌شوند؛ سپس اعضای کروموزوم فرزند، تک‌به‌تک و به‌صورت کاملاً تصادفی از اعضای این دو کروموزوم والد تشکیل می‌شود و درنهایت جواب جدیدی متشکل از اعضای این دو کروموزوم والد به‌دست می‌آید.

عملگر جهش: در الگوریتم ژنتیک، جهش رویدادی است که به‌طور تصادفی و با احتمال کم اتفاق می‌افتد و ضمن تغییر عناصر کروموزوم، از همگرایی زودرس الگوریتم جلوگیری می‌کند. در این پژوهش به‌منظور افزایش کارایی الگوریتم ژنتیک از یک عملگر جهش ابتکاری استفاده شده است. این رویکرد نخستین بار برای مسئله یادگیری ماشینی پیشنهاد شد [۴۲] که در آن عملگر جهش فقط زمانی اعمال می‌شود که میزان شباهت کروموزوم‌ها در هر جمعیت به حد خاصی برسد. ضریب شباهت بین دو کروموزوم از رابطه ۶ برآورد می‌شود:

رابطه (۶)
$$SC_{ab} = \frac{\sum_{i=1}^{N} \partial(x_{ija}, x_{ijb})}{N}$$

در رابطه ۶، X_{ijb} و X_{ija} عدد موجود در ستون i و سطر j در کروموزوم a و b هستند.

رابطه (۷)
$$\partial(x_{ija}, x_{ijb}) = \begin{cases} 1 & if \quad x_{ija} = x_{ijb} \\ 0 & if \quad otherwise \end{cases}$$

متوسط ضریب شباهت بین کروموزوم‌های یک جمعیت توسط رابطه ۸ تعیین می‌شود که در آن N تعداد کروموزوم‌ها است.

رابطه (۸)
$$\overline{SC} = \frac{\sum_{a=1}^{N-1} \sum_{b=a+1}^{N} SC_{ab}}{\binom{N}{2}}$$

در هر نسل فقط درصورتی‌که اندازه $\overline{SC}$ از یک آستانه خاص (بر اساس آزمایش‌های مقدماتی و روش سعی و خطا تعیین می‌شود) بیشتر شود، عملگر جهش بر روی کروموزوم‌ها قابل‌اعمال است. در هر تکرار مقدار ضریب شباهت بررسی شده و در صورتی که کروموزوم‌ها بیش‌ازحد به

هم شبیه باشند، عملگر جهش اعمال خواهد شد. در صورت برقراری شرط مقدماتی عملگر جهش بر روی کروموزوم ها اعمال خواهد شد. به‌منظور اعمال جهش در این پژوهش، ابتدا بهترین جواب از مجموعه جواب به‌عنوان کروموزوم فرزند انتخاب می‌شود و مقدار اعضای آن بر اساس قدم‌های زیر تغییر می‌کند:

۱. با استفاده از رابطه 0.1-0.1*it/Max_it، عدد p را تولید کنید. در این رابطه it شماره نسل و Max_it، حداکثر تعداد نسل است؛

۲. ژن سطر اول و ستون اول را انتخاب کرده و مقدار آن را با P*(2*RAND-1) جمع کنید؛

۳. مرحله دوم را برای تمامی ژن‌ها تکرار کنید.

شرط توقف الگوریتم: با توجه به اینکه الگوریتم ژنتیک بر پایه تولید و آزمون استوار است، نمی‌توان گفت که کدام یک از جواب‌های تولیدشده جواب بهینه است تا بر این اساس بتوان شرط توقف را پیداشدن جواب بهینه در جمعیت تعریف کرد؛ بنابراین الگوریتم از معیار تعداد تکرار به‌عنوان شرط توقف استفاده می‌کند که اندازه آن بر اساس تنظیم پارامتر تعیین خواهد شد.

٤. تحلیل داده‌ها و یافته‌های پژوهش

– **تنظیم پارامترها.** روش طراحی آزمایش‌های تاگوچی از روش‌های پرکاربرد در صنایع گوناگون است که معمولاً در کنترل‌کیفیت برون‌خطی کاربرد دارد [۳۵]. تاگوچی، کنترل‌کیفیت برون‌خطی را به سه مرحله طراحی سیستم، تنظیم پارامتر و طراحی تولرانس تقسیم‌بندی کرده است. مسئله اصلی این بررسی تنظیم پارامتر است که نخستین مرحله آن، تعیین عوامل موردبررسی و سطوح آن‌ها است. این عوامل همان پارامترهای ورودی الگوریتم ژنتیک هستند که مقادیر مربوط به هر سطح آن‌ها در جدول ۱، آمده است.

جدول ۱. پارامترها و سطوح آن‌ها

الگوریتم	پارامترها	تشریح پارامترها	سطح اول	سطح دوم	سطح سوم
GA	n-Pop	اندازه جمعیت	۱۰۰	۱۵۰	۲۰۰
	P_c	درصد تقاطع	۰/۷۵	۰/۸	۰/۸۵
	P_m	درصد جهش	۰/۰۳	۰/۰۵	۰/۰۸
	Max-it	تعداد تکرار	۵۰	۱۰۰	۱۵۰

ازآنجاکه الگوریتم ژنتیک دارای ۳ سطح و ۴ عامل است، تعداد ترکیب‌های ممکن شامل ۸۱ حالت خواهد بود؛ ولی با کاربرد روش تاگوچی فقط ۲۷ حالت ترکیبی وجود خواهد داشت که نشان‌دهنده کارایی بالای این روش است. متغیر پاسخ درنظرگرفته‌شده در هر آزمایش، مقدار تابع برازندگی است. جدول ۲، آزمایش‌های موردنظر و نتایج را برای الگوریتم ژنتیک نشان می‌دهد.

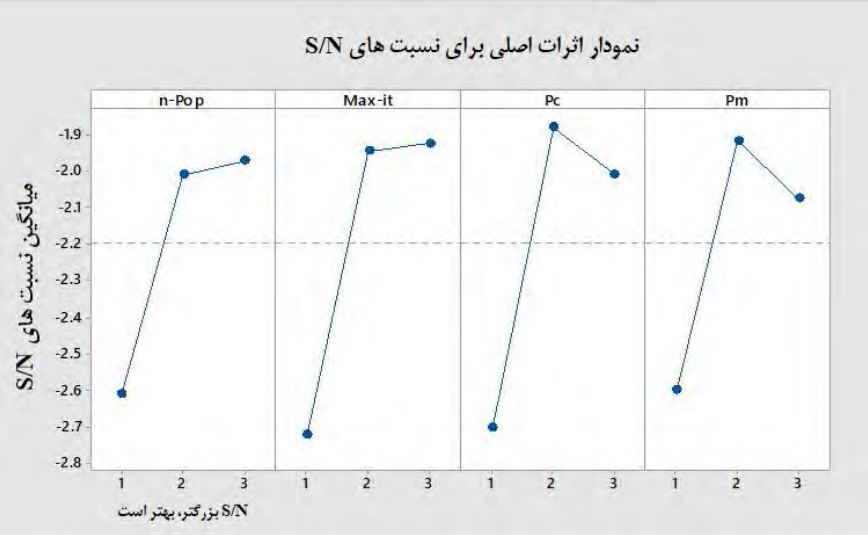
جدول ۲. آزمایش های تاگوچی و نتایج

آزمایش	n-Pop	Max-it	P_c	P_m	HGA	آزمایش	n-Pop	Max-it	P_c	P_m	HGA
۱	۱	۱	۱	۱	۰/۶۹	۱۵	۲	۲	۳	۱	۰/۸۳
۲	۱	۱	۱	۱	۰/۶۸	۱۶	۲	۳	۱	۲	۰/۸۰
۳	۱	۱	۱	۱	۰/۵۵	۱۷	۲	۳	۱	۲	۰/۸۱
۴	۱	۲	۲	۲	۰/۸۵	۱۸	۲	۳	۱	۲	۰/۷۸
۵	۱	۲	۲	۲	۰/۷۸	۱۹	۳	۱	۳	۲	۰/۸۰
۶	۱	۲	۲	۲	۰/۸۲	۲۰	۳	۱	۳	۲	۰/۷۸
۷	۱	۳	۳	۳	۰/۷۶	۲۱	۳	۱	۳	۲	۰/۷۹
۸	۱	۳	۳	۳	۰/۸۱	۲۲	۳	۲	۱	۳	۰/۷۹
۹	۱	۳	۳	۳	۰/۸۱	۲۳	۳	۲	۱	۳	۰/۸۴
۱۰	۲	۱	۲	۳	۰/۷۸	۲۴	۳	۲	۱	۳	۰/۷۳
۱۱	۲	۱	۲	۳	۰/۷۹	۲۵	۳	۳	۲	۱	۰/۸۷
۱۲	۲	۱	۲	۳	۰/۷۸	۲۶	۳	۳	۲	۱	۰/۸۰
۱۳	۲	۲	۳	۱	۰/۷۶	۲۷	۳	۳	۲	۱	۰/۷۸
۱۴	۲	۲	۳	۱	۰/۸۱						

سپس یک میانگین و یک نسبت S/N برای هر آزمایش محاسبه می‌شود. به‌طورکلی سه نوع نسبت S/N در روش طراحی تاگوچی وجود دارد که با توجه به مطلوبیت بیشتر مقادیر بالای متغیر پاسخ از نسبت S/N زیر استفاده می‌شود:

$$S/N = -10\,log\left(\frac{1}{n}\right)\sum_{i}\left(1/{y_i}^2\right)$$ رابطه (۹)

زمانی‌که کلیه نسبت‌های S/N و میانگین پاسخ‌ها به‌ازای هر یک از آزمایش‌ها محاسبه شد، روش تاگوچی از یک رویکرد نموداری برای تجزیه‌وتحلیل داده‌ها استفاده می‌کند. در این رویکرد، نمودارهای متوسط نسبت‌های S/N و نیز متوسط میانگین پاسخ‌ها برای هر عامل و به‌ازای سطوح مختلف آن‌ها ترسیم می‌شوند. سطوح بهینه هر عامل جایی است که نمودار S/N حداکثر شود. در شکل ۳، نمودار نسبت S/N نشان داده شده است.



شکل ۳. چگونگی تغییر مقادیر شاخص S/N در سطوح مختلف الگوریتم GA

– **بررسی و تحلیل تجربی.** پس از شرح ساختار الگوریتم پیشنهادی، در این بخش کارایی آن با تحلیل نتایج به‌دست‌آمده از مسائل نمونه ارزیابی می‌شود. ازآنجاکه مبانی نظری فاقد مسئله مشابهی در رابطه با مدل پیشنهادی است، چندین مسئله نمونه به‌طور تصادفی ایجاد می‌شوند و برای ارزیابی کارایی الگوریتم‌ها به‌کار می‌روند. بر این اساس، ابتدا الگوریتم‌ها برای کارکرد در بهترین شرایط تنظیم شده و مقادیر بهینه پارامترهای آن‌ها تعیین می‌شود. درنهایت، پس از ثبت نتایج به‌دست‌آمده از اجرای الگوریتم‌ها، کارایی آن‌ها با استفاده از آزمون‌های آماری با یکدیگر مقایسه می‌شود. برای تنظیم پارامترهای الگوریتم از نرم‌افزار Minitab17 استفاده شده است.

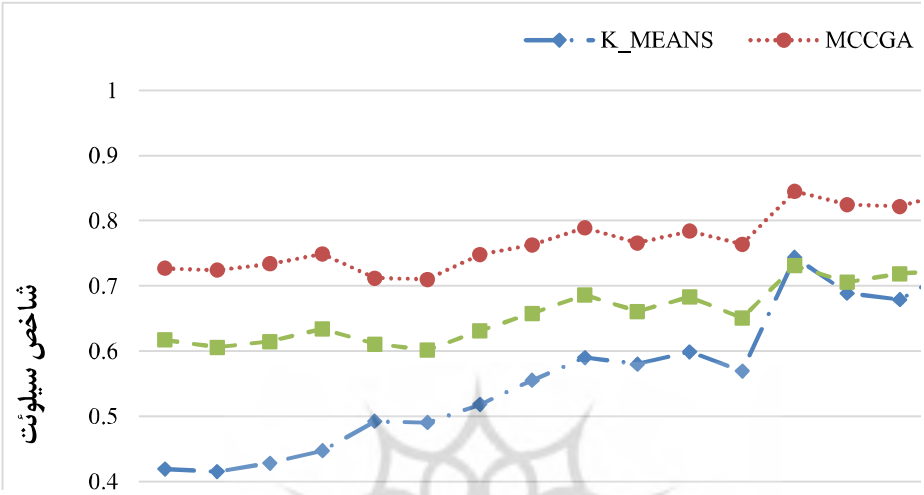
– **مقایسه کارایی الگوریتم‌ها.** به‌منظور بررسی و تحلیل نتایج، ۲۸ مسئله نمونه تصادفی اجرا می‌شود. مسائل تولیدشده با توجه به تعداد خوشه‌ها در ۷ دسته متفاوت قرار دارند. تمامی دسته‌ها شامل مسائلی با ۴۶۰ شرکت و ۱۸ شاخص برای هر شرکت هستند. با توجه به وزن شاخص‌ها، در هر گروه ۴ مسئله تصادفی تولید می‌شود. هر یک از مسائل نمونه ۵۰ بار تکرار شده و میانگین جواب‌های به‌دست آمده به‌عنوان نتیجه الگوریتم در نظر گرفته می‌شود؛ سپس عملکرد الگوریتم‌های پیشنهادی بر اساس معیار کیفیت جواب (کیفیت خوشه‌بندی) ارزیابی می‌شود. جدول ۳، نتایج محاسبات انجام‌شده را نشان می‌دهد. کارایی الگوریتم‌ها در شاخص کیفیت جواب در شکل ۴، نشان داده شده است. همان‌طور که مشخص است، جواب دو الگوریتم P2CLUST و MCCGA عملکرد بسیار بهتری نسبت به الگوریتم K– میانگین دارند.

کارایی الگوریتم‌ها بر اساس تعداد خوشه‌های اولیه نیز در شکل ۵، نشان داده شده‌است. با توجه به نتایج، بهترین عملکرد الگوریتم K– میانگین در حل مسائلی با ۶ خوشه است. با افزایش و کاهش تعداد خوشه‌ها، عملکرد K– میانگین به‌شدت کاهش می‌یابد که این امر نشان‌دهنده وابستگی شدید این الگوریتم به انتخاب تعداد خوشه‌های اولیه است. با این حال دو الگوریتم دیگر وابستگی کمتری به تعداد خوشه‌های اولیه دارند و در بیشتر مسائل نمونه عملکرد مناسبی از خود نشان می‌دهند. با توجه به شکل ۵، بالاترین کارایی الگوریتم‌های K– میانگین و MCCGA در شرایطی است که تعداد خوشه‌ها برابر ۶ باشد. نتایج الگوریتم P2CLUST نشان می‌دهد که تعداد ۷ خوشه ایده‌آل است؛ با این حال این الگوریتم با ۶ خوشه نیز نتایج مناسبی را ارائه می‌دهد.

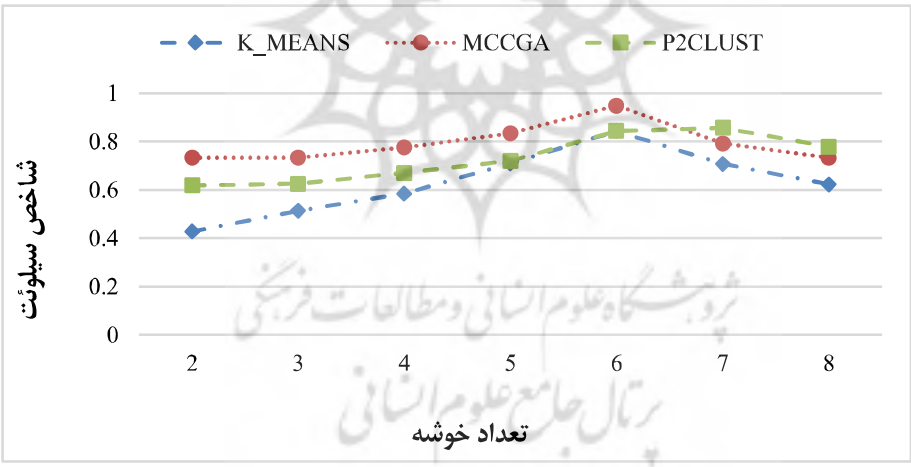
جدول ۳. نتایج محاسباتی الگوریتم‌ها در حل مسائل نمونه

شماره مسئله	شاخص سیلوئت			RPD		
	K-Means	MCCGA	P2CLUST	K-Means	MCCGA	P2CLUST
۱	۰/۴۱۹	۰/۷۲۷	۰/۶۱۷	۴۲/۳۶۶	۰/۰۰۰	۱۵/۱۳۱
۲	۰/۴۱۵	۰/۷۲۴	۰/۶۰۶	۴۲/۶۸۰	۰/۰۰۰	۱۶/۲۹۸
۳	۰/۴۲۸	۰/۷۳۴	۰/۶۱۵	۴۱/۶۸۹	۰/۰۰۰	۱۶/۲۱۳
۴	۰/۴۴۷	۰/۷۴۹	۰/۶۳۴	۴۰/۳۲۰	۰/۰۰۰	۱۵/۳۵۴
۵	۰/۴۹۲	۰/۷۱۲	۰/۶۱۱	۳۰/۸۹۹	۰/۰۰۰	۱۴/۱۸۵
۶	۰/۴۹۰	۰/۷۱۰	۰/۶۰۲	۳۰/۹۸۶	۰/۰۰۰	۱۵/۲۱۱
۷	۰/۵۱۸	۰/۷۴۸	۰/۶۳۱	۳۰/۷۴۹	۰/۰۰۰	۱۵/۶۴۲
۸	۰/۵۵۵	۰/۷۶۳	۰/۶۵۸	۲۷/۲۶۱	۰/۰۰۰	۱۳/۷۶۱
۹	۰/۵۹۰	۰/۷۸۹	۰/۶۸۶	۲۵/۲۲۲	۰/۰۰۰	۱۳/۰۵۴
۱۰	۰/۵۸۰	۰/۷۶۶	۰/۶۶۱	۲۴/۲۸۲	۰/۰۰۰	۱۳/۷۰۸
۱۱	۰/۵۹۹	۰/۷۸۴	۰/۶۸۳	۲۳/۵۹۷	۰/۰۰۰	۱۲/۸۸۳
۱۲	۰/۵۶۹	۰/۷۶۴	۰/۶۵۱	۲۵/۵۲۴	۰/۰۰۰	۱۴/۷۹۱
۱۳	۰/۷۴۴	۰/۸۴۵	۰/۷۳۱	۱۱/۹۵۳	۰/۰۰۰	۱۳/۴۹۱
۱۴	۰/۶۸۹	۰/۸۲۵	۰/۷۰۶	۱۶/۴۸۵	۰/۰۰۰	۱۴/۴۲۴
۱۵	۰/۶۷۹	۰/۸۲۲	۰/۷۱۹	۱۷/۳۹۷	۰/۰۰۰	۱۲/۵۳۰
۱۶	۰/۷۲۵	۰/۸۴۴	۰/۷۲۴	۱۴/۱۰۰	۰/۰۰۰	۱۴/۲۱۸
۱۷	۰/۸۴۵	۰/۹۵۵	۰/۸۳۹	۱۱/۵۱۸	۰/۰۰۰	۱۲/۱۴۷
۱۸	۰/۸۵۹	۰/۹۶۶	۰/۸۶۵	۱۱/۰۷۷	۰/۰۰۰	۱۰/۴۵۵
۱۹	۰/۸۲۸	۰/۹۳۱	۰/۸۳۰	۱۱/۰۶۳	۰/۰۰۰	۱۰/۸۴۹
۲۰	۰/۸۴۲	۰/۹۴۲	۰/۸۴۰	۱۰/۶۱۶	۰/۰۰۰	۱۰/۸۲۸
۲۱	۰/۷۱۲	۰/۸۰۲	۰/۸۴۴	۱۵/۶۴۰	۴/۹۷۶	۰/۰۰۰
۲۲	۰/۷۱۳	۰/۷۹۹	۰/۸۶۷	۱۷/۷۶۲	۷/۸۴۳	۰/۰۰۰

۲۳	۰/۷۲۹	۰/۸۰۹	۰/۸۵۷	۱۴/۹۳۶	۵/۶۰۱	۰/۰۰۰
۲۴	۰/۶۷۶	۰/۷۵۹	۰/۸۵۷	۲۱/۱۲۰	۱۱/۴۳۵	۰/۰۰۰
۲۵	۰/۶۱۱	۰/۷۲۲	۰/۷۸۴	۲۲/۰۶۶	۷/۹۰۸	۰/۰۰۰
۲۶	۰/۶۲۸	۰/۷۲۹	۰/۷۵۷	۱۷/۰۴۱	۳/۶۹۹	۰/۰۰۰
۲۷	۰/۶۵۰	۰/۷۷۱	۰/۷۹۸	۱۸/۵۴۶	۳/۳۸۳	۰/۰۰۰
۲۸	۰/۶۰۳	۰/۷۰۸	۰/۷۷۸	۲۲/۴۹۴	۸/۹۹۷	۰/۰۰۰



شکل ۴. کارایی الگوریتم‌ها در حل مسائل نمونه

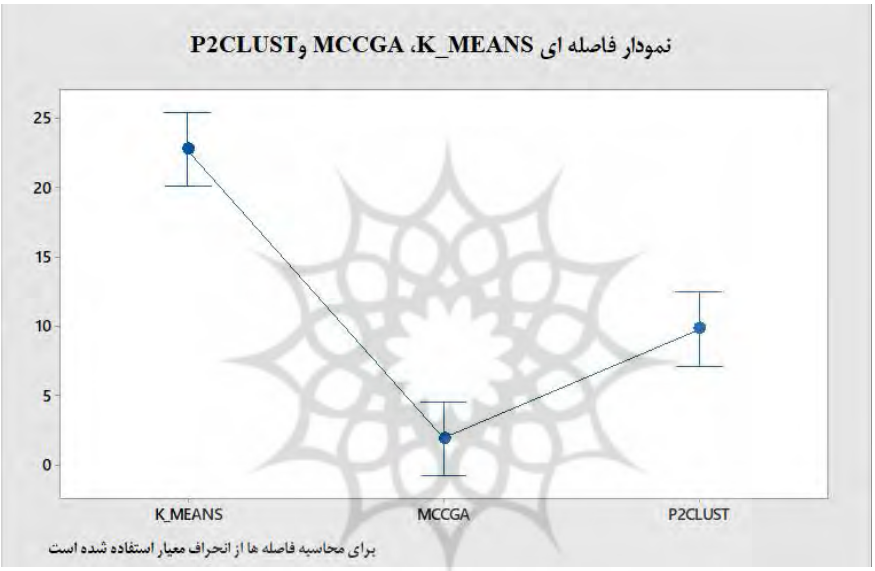


شکل ۵. کارایی الگوریتم‌ها بر اساس تعداد خوشه‌های اولیه

در راستای ارزیابی آماری الگوریتم‌ها از آزمون تحلیل واریانس استفاده شده است. به‌این‌منظور، پیش از انجام تحلیل واریانس، نتایج به‌دست آمده از الگوریتم‌های به‌کاررفته با استفاده از معیار درصد انحراف نسبی (رابطه ۱۰) نرمالایز شده است.

$$RPD_{ij} = \frac{Alg_{sol}(ij) - Best_{sol}(j)}{Best_{sol}(j)} \qquad \text{رابطه (۱۰)}$$

در رابطه ۱۰، i شماره الگوریتم، j شماره مسئله، $Best_{sol}(j)$ بهترین جواب به‌دست‌آمده در مسئله j و $Alg_{sol}(ij)$ جواب به‌دست‌آمده از الگوریتم i برای مسئله شماره j است. نتایج حاصل از آزمون تحلیل واریانس به شکل فواصل اطمینان ۹۵ درصد هم‌زمان توکی[1] در شکل ۶، نشان داده شده است.



شکل ۶. فاصله اطمینان ۹۵ درصد برای معیار کیفیت جواب در مسائل نمونه

با توجه به شکل ۶، می‌توان گفت که الگوریتم P2CLUST از نظر آماری بسیار بهتر از الگوریتم K– میانگین عمل می‌کند؛ ولی کارایی دو الگوریتم P2CLUST و MCCGA مشابه است. با این حال با توجه به اینکه بازه الگوریتم MCCGA از بازه الگوریتم P2CLUST پایین‌تر است، می‌توان ادعا کرد که الگوریتم MCCGA از عملکرد بهتری برخوردار است. درنهایت این مطالعه الگوریتم MCCGA را به‌عنوان الگوریتم کارآمدتری در خوشه‌بندی معرفی

1. Tukey

می‌کند؛ همچنین با توجه به اینکه بیشترین شاخص سیلوئت برای مسائل دارای ۶ خوشه است، می‌توان ادعا کرد که برای خوشه‌بندی مشتریان بانک با الگوریتم MCCGA، تعداد ۶ خوشه مناسب است و تعداد خوشه‌های کمتر یا بیشتر به کاهش کیفیت خوشه‌بندی منجر خواهد شد. تعداد بهینه خوشه‌ها در الگوریتم K- میانگین و P2CLUST نیز به‌ترتیب برابر ۶ و ۷ خوشه است؛ بنابراین تعداد ۶ خوشه برای مسئله موردبررسی، ایده‌آل‌ترین حالت ممکن است.

۵. نتیجه‌گیری و پیشنهادها

در این پژوهش، یک روش جدید برای خوشه‌بندی چندمعیاره متغیرها ارائه شد که سعی دارد نقایص روش مشهور P2CLUST که قبلاً به همین منظور توسعه یافته بود را برطرف کند. بدین منظور الگوریتم‌های K- میانگین و پرامتی با یکدیگر ترکیب شده و به‌جای محاسبه فواصل با توابع فاصله، از روابط باینری برای مقایسه دو متغیر استفاده شد. با توجه به پیچیدگی مسئله، الگوریتم ژنتیک برای بهینه‌سازی مسئله خوشه‌بندی به‌کار رفت که پارامترهای ژنتیک توسط روش تاگوچی تنظیم شدند. بر اساس نتایج می‌توان ادعا کرد که بهینه‌سازی الگوریتم خوشه‌بندی P2CLUST با استفاده از الگوریتم فراابتکاری مبتنی بر جمعیت ژنتیک برای حل مسئله خوشه‌بندی یک روش کارا است. به‌طورکلی الگوریتم‌های خوشه‌بندی نیازمند زمان محاسباتی بالایی هستند که نقطه ضعف اساسی آن‌ها به‌شمار می‌رود. در این مطالعه با هوشمندسازی عملگر جهش، الگوریتم ژنتیک ارتقا داده شده و الگوریتم ژنتیک ابتکاری معرفی شده است. نتایج محاسباتی نشان می‌دهد که الگوریتم توسعه‌یافته، قادر به دستیابی به جواب‌های بهتری است.

در راستای انجام پژوهش‌های آتی می‌توان به چندهدفه‌کردن مسئله با استفاده از سایر شاخص‌های ارزیابی عملکرد خوشه‌بندی که لزوماً با یکدیگر در تعارض هستند، اشاره کرد. استفاده از تکنیک پرامتی در ترکیب با سایر روش‌های خوشه‌بندی همچون خوشه‌بندی سلسله‌مراتبی به‌منظور بهبود عملکرد و استخراج پارامترهای روش‌های جدید طبقه‌بندی همچون پرامتی TRI از مواردی است که برای پژوهش‌های آتی پیشنهاد می‌شود.

منابع

1. Ahmadzadehgoli, N., Behzadi, M. H., & Mohammadpour, A. (2018). Clustering with Intelligent Linex K-means. *New Researches in Mathematics*, *14,* 5-14 (In Persian).
2. Aliheidari Bioki, T., & Khademi Zare, H. (2015). Improvement of DEA approach for clustering credit rating of customer in banks. *Modeling in Engineering, 41,* 59-74 (In Persian).
3. Alizadeh A, & Pooya A. (2017). Evaluating and clustering the Iranian banks and financial institutions based on website traffic indicators. *Organizational Resource Management Researches, 7(1),* 189-206 (In Persian).
4. Almeida, A. T., & Vetschera, R. (2012) A PROMETHEE-based approach to portfolio selection problems. *Computers & Operations Research, 39(5)*, 1010-1020.
5. Asgharizadeh, E. A., Bitaraf, A., & Ajeli, M. (2011). Developing a hybrid model using fuzzy PROMETHEE and multi-objective linear planning for outsourcing of warranty services, *Industrial Management Perspective, 2(2),* 43-60 (In Persian).
6. Azizi, S., & Balaghi Inanlou, M. H., (2016). Segmentation of Mobile Banking Users Based on Expectations: A Clustering Technique. *Production and Operations Management, 7(2),* 217-234 (In Persian).
7. Baroudi, R., Safia, N.B. (2010). Towards multicriteria analysis: a new clustering approach, in: Proceedings of the 2010 International Conference on Machine and Web Intelligence, pp. 126–131.
8. Brans, J. P., & Mareschal, B. (2005). PROMETHEE methods. In *Multiple criteria decision analysis: state of the art surveys*. (163-186). New York, New York: Springer.
9. Brans, J. P., & Vincke, P. (1985). PROMETHEE method for multiple criteria decision making. *Management Science, 31*, 647-656.
10. Capó, M., Pérez, A., & Lozano, J. A. (2017). An efficient approximation to the K-means clustering for massive data. *Knowledge-Based Systems, 117,* 56-69.
11. Cavalcante, C. A. V., Ferreira, R. J. P., & de Almeida, A. T. (2010). A preventive maintenance decision model based on multi-criteria method PROMETHEE II integrated with Bayesian approach. *IMA Journal of Management Mathematics, 21(4)*, 333-348.
12. Costa, C. B., De Corte, J. M. & Vansnick, J. C. (2005). On the mathematical foundations of MACBETH, in *Multiple Criteria Decision Analysis: state of the art surveys*, *Springer, 78,* 409-437.
13. De Smet, Y. (2013). P2CLUST: An extension of PROMETHEE II for multi-criteria ordered clustering. (2013). *IEEE International Conference on Industrial Engineering and Engineering Management*. Bangkok, Thailand.
14. De Smet, Y., & Guzmán, L. M. (2004). Toward multi-criteria clustering: An extension of the k-means algorithm. *European Journal of Operational Research, 158(2),* 390-398.
15. De Smet, Y., Nemery, P., & Selvaraj, R. (2012). An exact algorithm for the multi-criteria ordered clustering problem. *Omega,* 40(6), 861-869.
16. Dyer, J. S. (2005). Multi-Attribute Utility Theory (MAUT). In *Multiple Criteria Decision Analysis: State of the Art*. (265-292). New York, New York: Springer.
17. Faezi-Rad, M. A., & Pooya, A., (2016). Clustering of Online Stores from Supplier's point of view: Using Clusters Number Optimization in Two-Level SOM. *Industrial Management Studies, 34,* 905-943 (In Persian).

18. Fernandez, E., Navarro, J., & Bernal, S. (2010). Handling multicriteria preferences in cluster analysis, *European Journal of Operational Research, 202(3),* 819–827.
19. Figueira, J., Mousseau, V., & Roy, B. (2005). ELECTRE Methods. *Multiple Criteria Decision Analysis: State of the Art Surveys*. In *Multiple Criteria Decision Analysis: State of the Art*. (133-153). New York, New York: Springer.
20. Ghorbanpour A, Tallai G, & Panahi M. (2015). Clustering Customers of Refah Bank Branches Using Combination of Genetic Algorithm and C- Means in Fuzzy Environment. *Organizational Resources Management Researches, 5(3),* 153-168 (In Persian).
21. Hamedi, P., Khadivar, A., & Razmi, Z. (2013). Customer clustering for appointing rebating strategies, case study: Kadbano Co. *New Marketing Research, 3(3),* 135-150 (In Persian).
22. Han, J., Kamber, M., & Pei, J., (2011). Data Mining: Concepts and Techniques. 3rd edition, *Morgan Kaufmann*.
23. Holland, J. H. (1975). Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence. *U Michigan Press*.
24. Imani, A., & Abbasi, M. (2017). Customers Clustering Based on RFM Model by Using Fuzzy C-means Algorithm (Case Study: Zahedan City Refah Chain Store). *Public Management Researches, 37,* 251-276 (In Persian).
25. Islam, M. Z., Estivill-Castro, V., & Rahman, M. A., Bossomaier, T. (2018). Combining K-Means and a genetic algorithm through a novel arrangement of genetic operators for high quality clustering, *Expert Systems with Applications, 91,* 402-417.
26. Jafarnejad, A., Mohseni, M., & Abdollahi, A. (2014). Developing a hybrid fuzzy PROMETHEE-AHP approach for performance evaluation of the service supply chain (Case Study: Hospitality Industry), *Industrial Management Perspective, 4(2),* (In Persian).
27. Keshavarz Hadadha, A., Jalili Bal, Z., & Haji Yakhcha, S. (2018). Multi Criteria Decision Making Techniques and Knapsack Approach for Clustering, Evaluating and Selecting Projects. *Industrial Management Studies, 50,* 229-255 (In Persian).
28. Khadivar, A., & Hamedi, P. (2015). Providing Synthetic Data Mining Model Using Association Rules and Clustering for Determining Discounting Strategy (Case Study: Pegah Distribution Co). *Business Strategies, 5(3),* 39-52 (In Persian).
29. Khadivar, A., & Mojibian, F. (2018). Workshops Clustering Using a Combination Approach of Data Mining and MCDM. *Modern Researches in Decision Making, 3(2),* 107-128 (In Persian).
30. Krink, T., Paterlini, S., & Resti, A. (2007). Using differential evolution to improve the accuracy of bank rating systems. *Computational Statistics & Data Analysis*, *52(1),* 68-87.
31. Kumar, S., (2018). Optimal cluster analysis using hybrid K-Means and Ant Lion Optimizer. *Karbala International Journal of Modern Science* (In press).
32. Lai, R. K., Fan, C. Y., Huang, W. H., & Chang, P. C. (2018). Evolving and clustering fuzzy decision tree for financial time series data forecasting. *Expert Systems with Applications, 36 (2),* 3761-3773.
33. Liao, S. H., Ho, H. H., & Lin, H. W. (2008). Mining stock category association and cluster on Taiwan stock market. *Expert Systems with Applications, 35(1),* 19-28.

34. Meyer, P., & Olteanu, A. L. (2013). Formalizing and solving the problem of clustering in MCDA, *European Journal of Operational Research,* 227(3), 494-502.
35. Montgomery, D. C. (2009). Design and Analysis of Experiments. 8th Edition, *Wiley & Sons, Inc*.
36. Nabiloo, M., & Daneshpour, N. (2017). A clustering algorithm for categorical data with combining measures. *Soft Computing Journal*, *5(1),* 14-25 (In Persian).
37. Olson, D., & Zoubi, A. T. (2008). Using accounting ratios to distinguish between Islamic and conventional banks in the GCC region. *The International Journal of Accounting, 43(1),* 45-65.
38. Omidi, M., Razavi, H., & Mahpeykar, M. R. (2011). Selection of project team members based on the effectiveness criteria and PROMETHEE method, *Industrial Management Perspective, 2(1),* 113-134 (In Persian).
39. Rocha, C., Dias, L. C., & Dimas, I. (2013). Multi-criteria classification with unknown categories: A clustering-sorting approach and an application to conflict management. *Journal of Multi-Criteria Decision Analysis, 20,* 13-27.
40. Saaty, T. L. (2005). The analytic hierarchy and analytic network processes for the measurement of intangible criteria and for decision-making. *Multiple Criteria Decision Analysis: State of the Art Surveys*. In *Multiple Criteria Decision Analysis: State of the Art*. (345-405). New York, New York; Springer.
41. Sarrazin, R., De Smet, Y., & Rosenfeld, J. (2018). An extension of PROMETHEE to interval clustering. *Omega, 80,* 12-21.
42. Seifoddini, H., & Wolfe, P. M. (1986). Application of the similarity coefficient method in group technology. *IIE Transactions, 18(3)*, 271–277.
43. Silva, V. B., Morais, D. C., & Almeida, A. T. (2010). A multi-criteria group decision model to support watershed committees in Brazil. *Water Resources Management, 24(14)*, 4075-4091.
44. Yaghini, M., & Vard, M. (2012). Automatic Clustering of Mixed Data Using Genetic Algorithm. *Industrial Engineering & Production Management*, *23(2),* 187-197 (In Persian).
45. Yang, F., Sun, T., & Zhang, C., (2009). An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization. *Expert Systems with Applications, 36(9)*, 847-852 (In Persian).
46. Yu, S. S., Chu, S. W., Wang, C. M., Chan, Y. K., & Chang, T. C. (2017). Two improved *k*-means algorithms. *Applied Soft Computing, 68,* 747-755.
47. Zare Ahmadabadi, H., Rafiei Omam, M., & Naser Sadr Abadi, A. (2016). Market Clustering with Ant Colony Optimization (Comparative approach with k-means). *Business Administration research, 16,* 17-36 (In Persian).
48. Zhai, J., Cao, Y., Yao, Y., Ding, X., & Li, Y. (2017). Coarse and fine identification of collusive clique in financial market. *Expert Systems with Applications, 69,* 225-238.
49. Zhang, Y., Wang, C. D., Huang, D., Zheng, W. S., & Zhou, Y. R., (2018). TW-Co-*k*-means: Two-level weighted collaborative *k*-means for multi-view clustering. *Knowledge-Based Systems, 150,* 127-138.
50. Zhao, Y., Ming, Y., Liu, X., Zhu, E., Zhao, K., & Yin, J. (2018). Large-scale k-means clustering via variance reduction. *Neuro-computing*, *307*, 184-194.

Developing an Intelligent Multi Criteria Clustering Method Based on PROMETHEE

Amir Daneshvar*, Mahdi Homayounfar, Ania Farahmandnejad*****

Abstract

In recent years, a new issue called "multi-criteria clustering" has emerged that aims at grouping alternatives into homogeneous classes called clusters according to different evaluation criteria. Following the related studies in literature, by combining K-means algorithm and PROMETHEE technique, this paper aims to present a new multi-criteria clustering method. The parameters of the problem are the cluster separator profiles which genetic algorithm (GA) is used to optimize them. In the modeling process in each stage of updating responses, alternatives allocate to the nearest cluster according to the distance of their pure flow of privileges from the profiles. The mutation operator is only applied when the chromosomes' similarity level in each population reaches to a certain level which this intelligence reduces the computation time. Finally, by simulating the proposed algorithm and some well-known clustering algorithms based on the several financial databases the efficiency of the algorithm compared to other algorithms. The results show the algorithm, in addition to determine the optimal number of clusters in comparison to other algorithms, also provides better results.

Keywords: Multi-Criteria Clustering; Genetic Algorithm; K-Means Algorithm; Silhouette Index; PROMETHEE.

Received: Dec. 15, 2018, Accepted: Dec. 29, 2019.

* Assistant Professor in Industrial Management, Electronic Branch, Islamic Azad University, Tehran, Iran.

** Assistant Professor in Industrial Management, Rasht Branch, Islamic Azad University, Rasht, Iran (Corresponding Author).
E-mail: homayounfar@iaurasht.ac.ir

*** M.A. in Management of Information Technology, Electronic Branch, Islamic Azad University, Tehran, Iran.