DOI :10.22034/IAES.2019.101251.1665

اقتصاد کشاورزی/جلد ۱۳/شماره۲/صفحه های ۵۰-۳۳

مقایسه کاربرد الگوریتم‌های فراابتکاری کلونی مورچه‌ها و سلسله مراتبی در خوشه‌بندی شرکت‌های دانش‌بنیان کشاورزی

الهام مهرپرورحسینی، حامد رفیعی ، سید صفدر حسینی، امیرحسین چیذری[۱]

تاریخ دریافت: ۱۳۹۷/۱۰/۱۶ تاریخ پذیرش: ۱۳۹۸/۰۴/۲۴

چکیده

خوشه‌بندی یکی از عملیات مهم در داده‌کاوی به‌شمار می‌رود که نتایج آن برای محققان و سیاست‌گذاران در زمینه‌های مختلف به منظور تحلیل و برنامه‌ریزی مفید است. از آنجا که در سال‌های اخیر، اقتصاد دانش‌بنیان با حمایت از شرکت‌های دانش‌بنیان در ایران در حال توسعه است، تحلیل ویژگی‌های این شرکت‌ها و خوشه‌بندی آنها به منظور برنامه‌ریزی اثربخش، چشم‌انداز مناسبی برای سیاست‌گذاران فراهم می‌آورد. از این رو، در این مقاله شرکت‌های دانش‌بنیان کشاورزی ایران با استفاده از اطلاعات سال ۲۰۱۷ بر مبنای حوزه فناوری، تنوع محصولات و ارزش فروش محصول با استفاده از روش‌های خوشه‌بندی K-میانگین ساده و K-میانگین در ترکیب با الگوریتم‌های سلسله‌مراتبی و کلونی مورچه‌ها در نرم‌افزار MATLAB 16a خوشه‌بندی شدند. بنا بر نتایج، خوشه‌بندی K-میانگین در ترکیب با الگوریتم کلونی مورچه‌ها در مقایسه با دو روش دیگر توزیع متوازن‌تری از شرکت‌ها بین شش خوشه داشته است و میانگین ضریب نیمرخ ۰/۷ اعتبار این خوشه‌بندی را تأیید می‌کند. پیشنهاد می‌شود، به منظور اثربخشی و پرهیز از هدررفت بودجه حمایت از شرکت‌های دانش‌بنیان، برای شرکت‌هایی که به دلیل ویژگی‌های مشابه در یک خوشه قرار گرفته‌اند، برنامه‌های حمایتی همسانی طرح‌ریزی شود. برای شرکت‌های دانش‌بنیان با میانگین فروش اندک که در یک خوشه قرار گرفته‌اند برنامه‌های حمایتی ورود به بازار و برای شرکت‌های با تنوع محصول و درآمد متوسط بالا برنامه‌های توسعه بازار صادراتی توصیه می‌شود.

طبقه‌بندی JEL: C81, L2, M13, O31, Q16.

واژه‌های کلیدی: اقتصاد دانش‌بنیان، داده‌کاوی، خوشه‌بندی K-میانگین، الگوریتم سلسله مراتبی، الگوریتم کلونی مورچه‌ها.

مقدمه

[۱] به ترتیب:دانشجوی دکتری ،استادیار(نویسنده مسئول)، استاد، استادیار اقتصاد کشاورزی، دانشگاه تهران
Email: hamedrafiee@ut.ac.ir

در دهۀ 1970 میلادی با پیشرفت سریع فناوری‌های برتر در جهان، به ویژه در زمینه‌های ارتباطات و رایانه، الگوی رشد اقتصادی جهان به طور اساسی تغییر کرد و در پی آن، از دهۀ 1970 میلادی، دانش به عنوان مهم‌ترین سرمایه، جایگزین سرمایه‌های پولی و فیزیکی شد (Chen & et al., 2004). دانش از منبع‌های اصلی ایجاد اشتغال و ثروت و عامل دستیابی به توان رقابتی در بازارهای بین‌المللی در عصر جهانی شدن است که فرایندهای توسعۀ اقتصادی را عمق و شتاب می‌بخشد. تأثیر چشمگیر دانش بر عملکرد کلان اقتصادی کشورهای پیشرفته در دهه‌های گذشته، موجب پیدایش اصطلاح اقتصاد دانش‌بنیان شده است (Salmani، 2015). با وجود تعاریف پرشمار در مجامع اقتصادی یک تعریف عمده از اقتصاد دانش بنیان وجود دارد. براساس تعریف سازمان توسعه و همکاری‌های اقتصادی (2007)، اقتصاد دانش‌بنیان اقتصادی است که در آن دانش به عنوان موتور اصلی رشد اقتصادی عمل می‌کند و تولید و توزیع، مبتنی بر آن شکل گرفته و سرمایه‌گذاری در صنایع با محوریت دانش مورد توجه خاصی قرار گرفته است. سرمایه‌گذاری در دانش نیز عبارت است از مجموع هزینه‌های انجام شده در واحدهای تحقیق و توسعه و هر هزینه‌ای که بابت فعالیت‌های ارتقای دانش انجام گیرد.

در این راستا، کشورهای در حال توسعه برای دستیابی به هدف‌هایی مانند رشد اقتصادی و بهبود مزیت رقابتی در بازارهای جهانی، اقتصاد دانش‌بنیان را در دستور کار قرار داده‌اند، اما آنها در تبدیل دانش به ثروت با تنگناهای مختلفی روبه‌رو هستند. یکی از نارسایی‌هایی که در زمینه حمایت از شرکت‌های دانش‌بنیان در ایران وجود دارد، استفاده از ابزارهای حمایتی مشابه بدون در نظر گرفتن تفاوت ویژگی‌های شرکت‌ها است که موجب کاهش انگیزه و کارایی شرکت‌ها و هدررفت منابع می‌شود. بنا بر گزارش صندوق نوآوری و شکوفایی از سال ۲۰۱۳ تا ۲۰۱۶، بیش از چهار هزار میلیارد ریال به به طرح‌ها و هفت هزار میلیارد ریال در قالب انواع حمایت از شرکت‌های دارای گواهی دانش‌بنیان از سوی صندوق نوآوری و شکوفایی ریاست جمهوری اختصاص یافته است که براساس آیین‌نامه حمایت‌ها تفاوت چندانی از نظر تفاوت شرایط و ویژگی‌های شرکت‌ها در بهره‌مندی از آنها مشاهده نمی‌شود. این در حالی است که شرکت‌ها در حوزه‌های مختلف، تنوع محصولات، سطح فروش و دیگر ویژگی‌ها برای توسعه بازار و بهبود بهره‌وری نیازمند سطوح و انواع متفاوتی از حمایت‌ها هستند. از این رو دسته‌بندی شرکت‌های دانش‌بنیان با رویکردی چندجانبه‌نگر، با در نظر گرفتن ویژگی‌های مختلف آنها می‌تواند چشم‌انداز مناسبی را برای سیاست‌گذاران به منظور طرح‌ریزی ابزار حمایتی متناسب هر گروه

از شرکت‌ها فراهم آورد تا شرکت‌ها با بهره‌مندی از حمایت‌های ویژۀ خود از کارایی و قدرت رقابت‌پذیری بیشتری برخوردار شوند. این موضوع از آنجا که سازوکار حمایت از کسب و کارهای دانش‌بنیان در کشور نوپاست دارای اهمیت بیشتری است. زیرا شکل‌دهی و اصلاح سازوکارها در گام‌های نخست با سهولت بیشتری همراه است.

براساس آیین‌نامۀ حمایت از شرکت‌های دانش‌بنیان مصوب 2012 هیئت وزیران، شرکت‌ها در صورت دارا بودن شرایط ذکرشده در قانون، مجوز دانش‌بنیانی را از معاونت علمی و فناوری ریاست جمهوری دریافت می‌کنند و همۀ مراحل اجرایی این قانون توسط این معاونت انجام می‌شود. در این راستا، معاونت علمی ریاست جمهوری از سال 2012 به صورت دوره‌ای اقدام به انتشار فهرست محصولات دانش‌بنیان مورد پذیرش در 14 حوزه می‌کند که در سال 2017 تعداد کل شرکت‌های دانش‌بنیان ثبت شده 2797 بوده است. از این میان 102 شرکت در حوزه زیست‌فناوری به تولید 272 محصول در زمینه کشاورزی، دامی و گیاهی پرداخته‌اند (Vice Presidency for Science and Technology of the Islamic Republic of Iran, 2017). با توجه به تأکید سند چشم‌انداز 2025 کشور و بند اول سیاست‌های اقتصاد مقاومتی بر تقویت اقتصاد دانش‌بنیان و همچنین اهمیت نفوذ دانش در بخش کشاورزی به منظور تحول آن از یک بخش سنتی به پیشرفته، بررسی و برنامه‌ریزی برای شرکت‌های دانش‌بنیان فعال در زمینه کشاورزی، دامی و گیاهی از اهمیت بیشتری برخوردار است. از این رو، در این مقاله به ارزیابی و خوشه‌بندی شرکت‌های دانش‌بنیان کشاورزی بر مبنای شاخص‌های درآمد ناشی از فروش محصولات دانش‌بنیان هر شرکت، تنوع محصولات و زیردسته فناوری آنها با استفاده از روش K-میانگین ساده و الگوریتم‌های فراابتکاری ترکیبی با K-میانگین و مقایسه این روش‌ها به منظور دستیابی به مناسب‌ترین خوشه‌بندی پرداخته می‌شود. در ادامه به نتایج برخی مطالعات مربوط به خوشه‌بندی در علوم اقتصادی اشاره خواهد شد.

امروزه خوشه‌بندی از روش‌های رایج در مطالعات علوم مختلف از جمله اقتصاد به شمار می‌رود. Linares & et al. (2018) با استفاده از نسبت‌های مالی استاندارد شده، شرکت‌های عضو بورس اوراق بهادار را با روش‌های مختلف محاسبه فاصله خوشه‌بندی کردند. به طوری که شرکت‌های با ساختار مالی همانند در یک گروه قرار گرفتند. Rahman & et al. (2017) مشتریان یک شرکت بیمه‌ای در بنگلادش را برای سال‌های ۲۰۱۴-۲۰۱۱ با روش‌های مختلف

خوشه‌بندی کردند. Zare & et al. (2017) بازار هدف یک کارخانه کاشی را به طور استانی با دو روش K-میانگین و الگوریتم کلونی مورچه‌ها خوشه‌بندی کردند که براساس شاخص اعتبار خوشه‌بندی، روش الگوریتم کلونی مورچه‌ها نتایج بهتری داشت. Daei & Li (2016) روش سلسله مراتبی را برای خوشه‌بندی در زنجیره ارزش با در نظر گرفتن تنوع محصولات به کار گرفتند. Aliheidari & Khademizare (2015) به خوشه‌بندی مشتریان اعتباری بانک‌ها براساس نسبت‌های مالی با استفاده از روش تحلیل پوششی داده‌ها پرداختند. در بررسی Shirkavand(2014) برای خوشه‌بندی مشتریان صنعت بیمه، نخست 30 شاخص به عنوان ارزش‌های مورد انتظار مشتریان این صنعت انتخاب و در قالب پرسشنامه با پرسش‌های دارای طیف لیکرت دیدگاه‌های مشتریان جمع‌آوری شد. آنگاه با استفاده از روش K-میانگین مشتریان صنعت به چهار خوشهٔ قیمت‌گرا، خدمات‌گرا، سهولت‌گرا و رابطه‌گرا تقسیم شدند. در نهایت نتایج آزمون کای دو پیرسون نشان داد، خوشه‌های مشتریان از نظر متغیرهای سن، تأهل، سطح تحصیلات و درآمد تفاوت معناداری با یکدیگر ندارند. Epetimehin (2011)، بخش‌بندی بازار را به عنوان ابزاری مؤثر در رضایت مشتریان در صنعت بیمه نیجریه بررسی کردند. Brandt (2005) با استفاده از روش‌های مختلف خوشه‌بندی به بخش‌بندی بازار نوشیدنی‌ها در کانادا پرداخته است و این کار را روشی برای دستیابی به استراتژی‌های بازاریابی مناسب دانسته است. Kuo & et al. (2005) با روش K-میانگین در ترکیب با الگوریتم کلونی مورچگان، بازار تلویزیون پلاسما در تایوان را بخش‌بندی کردند. Alfansi & Sargeant (2000) بازار مشتریان بانکی اندونزی را براساس ویژگی‌های جمعیت‌شناختی[1] و سود با استفاده از روش خوشه‌بندی، بخش‌بندی کردند. همان‌طور که در پژوهش‌های انجام گرفته با روش خوشه‌بندی مشاهده می‌شود، امروزه خوشه‌بندی یکی از روش‌های رایج در بررسی و دسته‌بندی مشاهده‌های یک پژوهش براساس ویژگی‌های آنهاست که در بررسی‌های اقتصادی از جمله بخش‌بندی[2] بازارهای هدف نیز گسترش پیدا کرده است. برای خوشه‌بندی از الگوریتم‌های مختلفی استفاده می‌شود که کارایی و دقت آنها بر مبنای نوع پراکنش داده‌ها می‌تواند متفاوت باشد. وجه تمایز این مقاله با دیگر پژوهش‌ها آن است که برای نخستین بار به بررسی شرکت‌های دانش‌بنیان با رویکرد خوشه‌بندی آنها پرداخته و برای این منظور از سه الگوریتم متفاوت استفاده و بنا بر نتایج مناسب‌ترین روش را انتخاب کرده است. در بخش بعدی

[1] Demogeraphic
[2] Segmentation

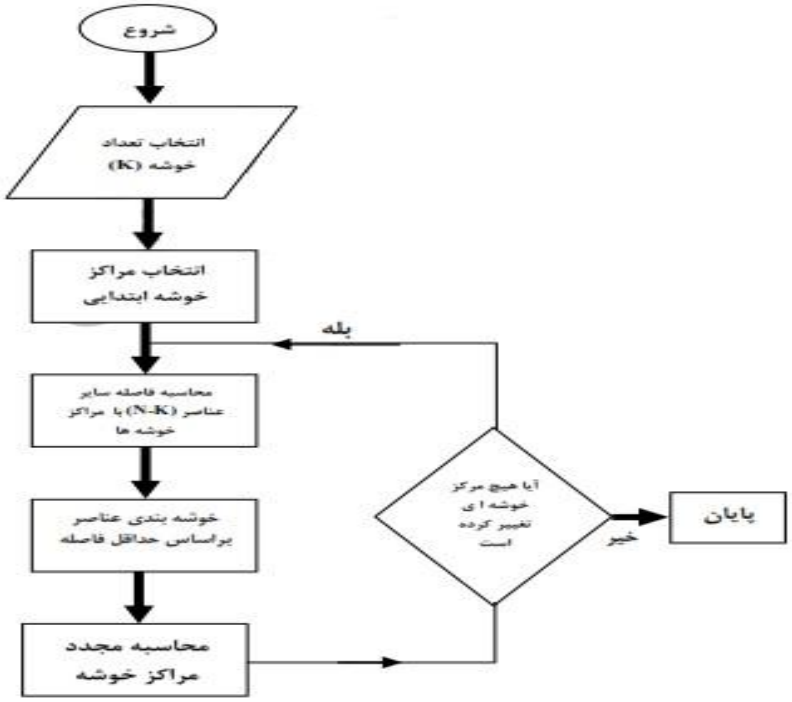
روش پژوهش با توجه به مطالعات انجام شده دیگر پژوهشگران و سپس نتایج و پیشنهادها بیان خواهد شد.

روش تحقیق

داده‌کاوی به فرایند استخراج الگوهای پنهان و یا ویژگی‌های جالب و سودمند از مجموعه داده‌ها گفته می‌شود که با استفاده از آن می‌توان به تصمیم‌گیری و پیش‌بینی رفتار آینده پرداخت. خوشه‌بندی در داده‌کاوی یکی از عملیات مهم در نتیجه‌گیری داده‌کاوی روی داده‌ها به‌شمار می‌آید. خوشه‌بندی، افرازبندی[1] یک گروه متنوع به تعدادی زیرگروه همسان یا گروه‌بندی مجموعه‌ای از اشیاء به کلاسی از اشیاء مشابه می‌باشد، در هر خوشه باید داده‌هایی شبیه به هم قرار گیرند و کمترین شباهت را با داده‌های موجود در دیگر خوشه‌ها دارا باشند (Linares & et al.,2018).

الگوریتم K-میانگین یکی از روش‌های خوشه‌بندی است که نخستین بار توسط Macqueen (1967) ارایه شد. این روش برای خوشه‌بندی داده‌هایی طراحی شده است که به صورت کمّی بوده و خوشه دارای مرکزی به نام میانگین است. در این روش، نخست اشیاء به صورت تصادفی به k خوشه تقسیم می‌شوند و در گام بعدی فاصلهٔ هر یک از اشیاء از مرکز خوشه خود محاسبه می‌شود. در صورتی که فاصلهٔ شیء مورد نظر از میانگین خوشهٔ خود زیاد و به خوشه دیگری نزدیک‌تر باشد، این شی به خوشه‌ای که نزدیک‌تر است اختصاص می‌یابد. این کار آن‌قدر تکرار می‌شود تا تابع خطا حداقل شود و یا اعضای خوشه تغییر نیابد (Sadeghian, 2013). در شکل (1) مراحل اجرای الگوریتم K-میانگین به طور خلاصه نشان داده شده است.

[1] segmentation



شکل (۱) الگوریتم خوشه‌بندی K-میانگین

Figure (1) K-means clustering algorithm

اگر D مجموعهٔ داده‌ها با n شیء باشد و C_i بیانگر خوشه‌ها به تعداد k خوشه باشد، در این صورت تابع خطا (EF) مجموع فاصله‌های هر شی از مرکز خوشهٔ خودش به شکل رابطه (1) تعریف می‌شود:

$$EF = \sum_{i=1}^{k} \sum_{X \in C_i} d(X, \mu(C_i)) \qquad (۱)$$

که در آن μ نشان‌دهندهٔ مرکز خوشه و $d(X,\mu(C_i))$ فاصلهٔ هر شیء از مرکز خوشه است که این فاصله می‌تواند بر پایهٔ اقلیدسی محاسبه شود. به دلیل آن که در خوشه‌بندی‌های مرکزگرا، تابع خطایی وجود دارد که لازم است حداقل شود، می‌توان به مسئله‌های خوشه‌بندی مرکزگرا به دید مسئله‌های بهینه‌سازی نگریست (Momeni, 2016).

از آنجا که در روش K-میانگین ساده، مراکز خوشه‌ها به صورت تصادفی انتخاب می‌شود، نتیجهٔ خوشه‌بندی وابسته به انتخاب مراکز اولیه است که از نقاط ضعف این روش به شمار می‌رود. برای اصلاح این موضوع می‌توان از الگوریتم‌های دیگری در ترکیب با K-میانگین استفاده کرد که با انتخاب مناسب‌ترین مراکز خوشه‌ها، تابع خطای فاصله رابطهٔ (1) را حداقل کند.

خوشه بندی سلسله مراتبی یکی از پرکاربردترین روش‌های خوشه‌بندی است. در این روش در آغاز فاصله دو به دوی مشاهدات از هم محاسبه می‌شود. پس از تعیین فاصله دو به دوی مشاهدات، با توجه به نزدیکی آنها نسبت به یکدیگر، با هم تشکیل یک خوشهٔ جدید می‌دهند. این کار تا آنجا ادامه پیدا می‌کند تا همه مشاهدات در یک خوشه قرار می‌گیرند (Daei & Li, 2016; Momeni, 2014). در این مقاله از این الگوریتم برای تعیین مراکز خوشه‌ها استفاده شده است. به طوری که ابتدا فاصله هر یک از مشاهدات به صورت دو به دو با دیگر مشاهدات اندازه‌گیری شده و برمبنای تعداد خوشه‌های تعیین شده، مشاهداتی که کمترین فاصله را با بیشترین تعداد مشاهدات دارند به عنوان مراکز آغازین انتخاب می‌شوند و دیگر مراحل مانند روش K-میانگین انجام می‌شود. در واقع در این روش تابع خطا برای همه حالت‌ها به عنوان مرکز انتخاب شدن همهٔ مشاهدات محاسبه می‌شود. این روش به دلیل محاسبه فاصله برای تک‌تک مشاهدات برای زمانی که تعداد مشاهدات بسیار زیاد است به دلیل حجم بالای محاسبات توصیه نمی‌شود.

روش دیگری که برای بهبود عملکرد خوشه‌بندی به روش K-میانگین در مطالعات پیشنهاد شده است، ترکیب آن با الگوریتم فراابتکاری کلونی مورچه‌ها است (Kuo & et al. 2005). الگوریتم‌های فراابتکاری به الگوریتم‌هایی گفته می‌شود که به‌طور معمول از طبیعت الهام گرفته شده‌اند و به طور متوسط بهترین تقابل کیفیت و زمان حل مسئله را به همراه دارند. اما به دلیل ساختار هوشمندانه، محاسبات زیادی دارند که بسیار زمان‌بر است (Golichenar & Saniee, 2012). از این رو، برای حل مشکل سرعت پایین و همگرایی دیررس این الگوریتم‌ها از ترکیب آنها با الگوریتم K-میانگین استفاده می‌شود که موجب حل مشکل همگرایی زودرس و بیرون آمدن از بهینهٔ محلی به بهینهٔ فراگیر در الگوریتم K-میانگین ساده می‌شود و نیاز به تعیین تعداد خوشه‌ها از آغاز کار نیز ندارد (Kuo & et al., 2005). مراحل این روش به شرح زیر است:

در اولین گام پارامترها، شامل تعداد خوشه‌های اولیه و مراکز اولیه با استفاده از الگوریتم مورچه‌ها تعیین می‌شود. در این روش هر مورچه (k) در یک گره از گراف مربوط به مسئله(i) قرار می‌گیرد و سپس با احتمال (P) به گرهٔ دیگری(j) حرکت کرده و به همین ترتیب ادامه می‌دهد تا یک پاسخ کامل ایجاد شود. در این روش مراکز خوشه‌ها بر مبنای مقدار فرومونی که مورچه‌ها هنگام حرکت در مسیر بین مشاهدات برای جای می‌گذارند تعیین می‌شود و فرض بر این است که در هر بار حرکت در مسیر مقدار فرومون ثابتی آزاد می‌شود (Zare & et al., 2017). در این حالت احتمال انتخاب مسیر ij که دو نقطه از خوشه‌های فرضی انتخاب شده در الگوریتم هستند، توسط مورچه k با رابطه (2) تعریف می‌شود:

$$P_{ij}^{k} = \frac{\tau_{ij}^{\alpha}\delta_{ij}^{\beta}}{\sum_{c}^{nc}\tau_{ic}^{\alpha}\delta_{ic}^{\beta}} \qquad (2)$$

که در آن nc تعداد خوشه‌ها و c شماره خوشه، τ مقدار فرومون، δ مشاهده‌پذیری(معکوس فاصله اقلیدسی بین i و j) و α و β به ترتیب اهمیت نسبی مسیر و مشاهده‌پذیری است. چنانچه P_{ij} بزرگتر از دیگران باشد، مورچه k از نقطه i به j حرکت خواهد کرد. به عبارت دیگر، شیء i متعلق به خوشه‌ای است که شیء j در آن قرار دارد. در گام بعدی مقدار فرومون به $\tau_{ij}+Q/TWVC$ افزایش می‌یابد که Q جزء ثابت و TWVC مجموع واریانس درونی خوشه است که با رابطه (3) محاسبه می‌شود:

$$TWVC = \sum_{k=1}^{nc}\sum_{i\in k}(O_i, O_{center}(T_k))^2 \qquad (3)$$

که در آن،O_i اعضا و $O_{center}(T)$ شیء میانی مجموعه اشیایی است که در خوشه T قرار گرفته‌اند و $k=1,2,3,\ldots,nc$ است. $O_{center}(T_k)$ نیز از رابطه (4) محاسبه می‌شود:

$$O_{center}(T) = \frac{1}{n_T}\sum_{O_i\in T} O_i \qquad (4)$$

که در آن n_T تعداد اعضای خوشه T است. پس از محاسبه TWVC اگر مقدار آن تغییر کرد محاسبات دوباره به مرحله نخست بازمی‌گردد و تعداد خوشه و مراکز جدید تعریف می‌شود و این عملیات تا جایی تکرار می‌شود که مجموع واریانس درونی خوشه‌ها ثابت بماند (Kuo & et al. 2005). بدین ترتیب تعداد و مراکز اولیه خوشه‌ها با استفاده از الگوریتم کلونی مورچه‌ها

مشخص می‌شود و ادامه مراحل که جای گرفتن اعضا درون خوشه‌هاست مانند الگوریتم K-میانگین در شکل (1) خواهد بود.

برای اعتبارسنجی خوشه‌بندی از شاخص ضریب نیمرخ[1] استفاده می‌شود که میانگین آن برای کل مشاهدات و برای هر خوشه محاسبه‌پذیر است. این شاخص میزان نزدیک شدن مشاهدات درون خوشه و میزان آزادی و فاصله آنها از خوشه همسایه را نشان می‌دهد و بیشتر بودن این شاخص بیانگر تعلق بیشتر اعضا به هر خوشه است. منفی بودن این مقدار به این معناست که مشاهده مربوطه به خوشه‌ای که در آن قرار گرفته تعلق ندارد (Momeni, 2016). با وجود این، مقدار ضریب نیمرخ مشاهدات و میانگین آن برای کل آنها تنها معیار انتخاب الگوریتم مناسب و تعداد خوشه‌ها نیست و توزیع مشاهدات درون خوشه‌ها نیز در انتخاب روش خوشه‌بندی دارای اهمیت است (Zare & et al., 2017).

در ادامه با استفاده از سه روش بیان شده، شرکت‌های دانش‌بنیان بخش کشاورزی بر مبنای سه ویژگی زیردستهٔ فناوری، تعداد محصولات دانش‌بنیان شرکت و مجموع مقدار فروش این محصولات در شرکت، به عنوان شاخصی برای نشان دادن اندازه بازار شرکت، خوشه‌بندی و بنا بر مقایسهٔ نتایج، کارآمدترین نوع خوشه‌بندی برای این داده‌ها با توجه به شاخص ضریب نیمرخ و توزیع شرکت‌ها در خوشه‌ها انتخاب خواهد شد. برای انجام خوشه‌بندی و اجرای الگوریتم‌ها از داده‌های معاونت علمی ریاست جمهوری در سال 2017 و بستهٔ نرم‌افزاری MATLAB 16a استفاده شده است.

نتایج و بحث

همان‌طور که پیش از این بیان شد، هدف این مقاله خوشه‌بندی شرکت‌های دانش‌بنیان بخش کشاورزی است، به صورتی که شرکت‌های دارای بیشترین وجوه اشتراک در یک گروه قرار بگیرند. برای این منظور، نخست از روش خوشه‌بندی K-میانگین ساده، از رایج‌ترین ابزار خوشه‌بندی در مطالعات مختلف، استفاده شد؛ اما از آنجا که یکی از نقاط ضعف این روش تصادفی بودن انتخاب مراکز خوشه‌هاست، از این رو، دو روش K-میانگین ترکیبی با الگوریتم‌های سلسله مراتبی و کلونی مورچه‌ها نیز به کار گرفته شدند که نتایج خوشه‌بندی مربوط به سه روش در جدول (1) قابل مشاهده است.

---

[1] Silhouette Coefficient

جدول (۱) مقایسه انواع روش‌های خوشه‌بندی و اعتبار آنها
Table (1) Comparison of Types of Clustering Methods and Their Validity

نوع خوشه‌بندی Type of clustering	شماره خوشه‌ها Number of clusters	1	2	3	4	5	6	7	مجموعه خوشه‌ها Total clusters
K-means ساده Simple K-means	تعداد شرکت‌ها Numbers of companies	91	1	1	3	2	3	1	102
	میانگین ضریب نیمرخ The average of the silhouette coefficient	0/99	1	1	0/85	1	1	1	0/99
K-means ترکیبی با الگوریتم سلسله مراتبی K-means combines with hierarchical algorithm	تعداد شرکت‌ها Numbers of companies	50	26	18	5	3			102
	میانگین ضریب نیمرخ The average of the silhouette coefficient	0/74	0/79	0/64	0/25	0/61			0/71
K-means ترکیبی با الگوریتم کلونی مورچه‌ها K-means combines with ant colony algorithm	تعداد شرکت‌ها Numbers of companies	33	25	18	4	4	18		102
	میانگین ضریب نیمرخ The average of the silhouette coefficient	0/61	0/76	0/75	0/56	0/74	0/76		0/70

منبع: یافته‌های تحقیق Source: Research finding

در روش K-میانگین ساده، براساس ضریب شاخص نیمرخ 99 درصدی، مناسب‌ترین تعداد خوشه‌ها 7 بود، در حالی‌که تعداد بهینهٔ خوشه‌ها در روش K-میانگین ترکیبی با الگوریتم سلسله مراتبی و کلونی مورچه‌ها به ترتیب 5 و 6 با ضرایب نیمرخ 71 و 70 درصد بود. با وجود آنکه، ضریب نیمرخ در روش K-میانگین ساده در مقایسه با دو روش دیگر بزرگ‌تر است، اما توزیع نامتوازن شرکت‌ها بین خوشه‌ها و وجود سه خوشهٔ تک‌عضوی موجب می‌شود، دو روش

دیگر که هم از نظر شاخص ضریب نیمرخ معتبر هستند و هم توزیع متوازن‌تری از شرکت‌ها در خوشه‌ها نشان می‌دهند، مناسب‌تر تشخیص داده شوند. در بین دو روش K-میانگین ترکیبی با الگوریتم سلسله مراتبی و کلونی مورچه‌ها نیز روش K-میانگین ترکیبی با الگوریتم کلونی مورچه‌ها که میانگین ضرایب نیمرخ در همه گروه‌های آن بیشتر از 50 درصد است و بیان می‌کند، اعضای هر خوشه بیشترین تعلق را به آن خوشه دارند و توزیع متوازن‌تری را نشان می‌دهد، انتخاب می‌شود.

همان‌طور که در جدول (2) مشاهده می‌شود، بر مبنای روش K-میانگین ترکیبی با کلونی مورچه‌ها، بیشترین تعداد شرکت‌ها و محصولات در خوشه یک و کمترین تعداد در خوشه‌های چهار و پنج که به ترتیب با 27293 و 5404 میلیون ریال دارای بیشترین میانگین فروش به ازای هر شرکت هستند قرار دارند. علت اندک بودن تعداد اعضای این خوشه‌ها آن است که تعداد کمی از شرکت‌ها دارای تعداد زیادی محصول دانش‌بنیان با فروش بالا هستند و اغلب شرکت‌ها با تعداد کمی محصول بازار کوچکی را کسب کرده‌اند که لازم است، شرایط این شرکت‌ها در استفاده از ابزار حمایتی، متفاوت با دیگر شرکت‌ها باشد. تنوع محصولات و فروش شرکت‌ها، انعطاف‌پذیری شرکت در بازارهای مختلف، اندازه بازار و توانایی شرکت در توسعه بازار را نشان می‌دهد. دلیل پایین بودن میانگین فروش در خوشهٔ سوم آن است که بیشتر اعضای این خوشه، شرکت‌های دانش‌بنیان نوپا هستند که هنگام دریافت گواهی دانش‌بنیان، محصول آنها به مرحلهٔ فروش نرسیده بوده است.

جدول (۲) وضعیت توزیع شرکت‌ها، محصولات و فروش آنها در خوشه‌بندی K-میانگین ترکیب با کلونی مورچه‌ها

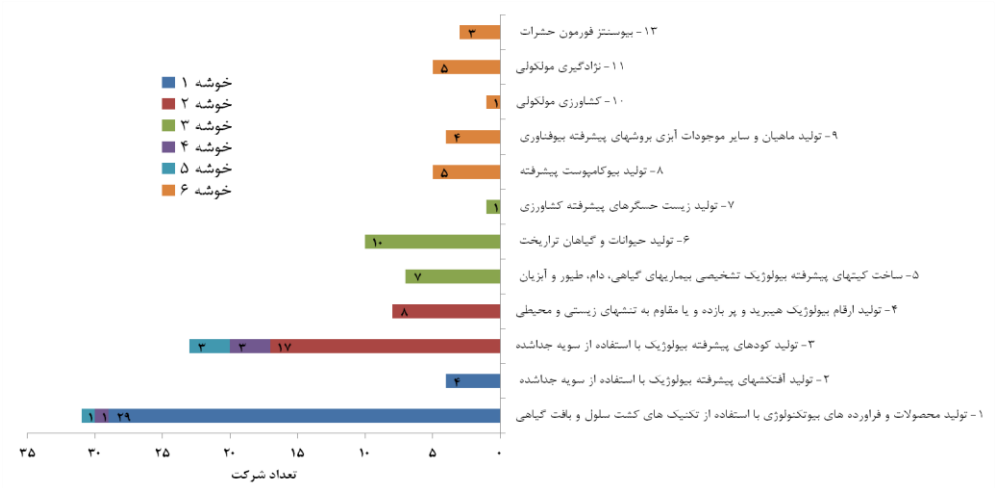
Table (2) Distribution status of companies, products and sales in the clustering K-means combined with Ant colony

مجموعه گروه‌ها Total clusters	6	5	4	3	2	1	شماره خوشه Number of clusters
102	18	4	4	18	25	33	Numbers of companies
272	38	54	48	30	37	65	تعداد محصولات Numbers of products
155719	9212	21615	109172	502	7864	7354	مجموع ارزش فروش (میلیون ریال) Total sale value (Million Rials)
1527	512	5404	27293	28	315	223	میانگین ارزش فروش به ازای هر شرکت (میلیون ریال) Average sale value per company (Million Rials)

منبع: یافته‌های تحقیق Source: Research finding

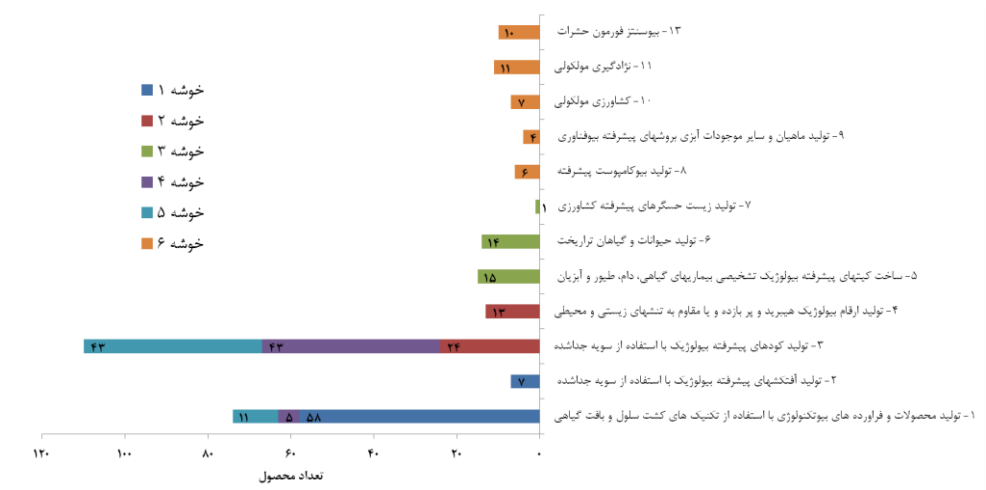
همان‌طور که بیان شد، براساس دسته‌بندی محصولات دانش‌بنیان در معاونت علمی و فناوری ریاست جمهوری، دستهٔ محصولات کشاورزی، دامی و گیاهی خود دارای سیزده زیردسته است که تاکنون در زیردستهٔ دوازده، یعنی تولید دارو و واکسن زیستی برای دام و طیور، محصولاتی ثبت نشده است. توزیع شرکت‌ها و محصولات بر مبنای خوشه‌بندی انجام شده، در هر یک از دوازده دستهٔ دارای محصول به ترتیب در نمودارهای (1) و (2) نشان داده شده است.

از 33 شرکت عضو خوشهٔ یک، 29 شرکت عضو زیردستهٔ یک، تولید محصولات و فراورده‌های زیست‌فناوری با استفاده از فناوری‌های کشت سلول و بافت گیاهی، هستند. بیشتر شرکت‌ها و محصولات عضو دسته‌های سه و چهار، تولید کودهای پیشرفتهٔ زیستی و ارقام زیستی هیبرید و پربازده، نیز به خوشهٔ دو تعلق دارند. خوشه‌های سه و شش به ترتیب با 30 و 38 محصول و 18 شرکت بیشترین پراکندگی را در زیردسته‌ها دارند. همهٔ شرکت‌های زیردسته‌های پنج تا هفت که اغلب شرکت‌های نوپا هستند، در خوشهٔ سه و زیردسته‌های هشت تا سیزده در خوشهٔ شش قرار دارند. اعضای خوشه‌های چهار و پنج که اغلب شرکت‌های پرسابقه با تنوع محصول و درآمد بالا هستند در زیردسته‌های اول و سوم قرار دارند. در دستهٔ کودهای پیشرفته، سه شرکت خوشه‌های چهار و پنج، هریک 43 محصول دارند. شایان یادآوری است، سابقه فعالیت شرکت‌ها نیز یکی از وجوه تمایز آنهاست که در سیاست‌گذاری در نظر گرفته نمی‌شود. با توجه به در دسترس نبودن اطلاعات سابقه همه شرکت‌های مورد بررسی، این ویژگی خود را در درآمد شرکت‌ها که در خوشه‌بندی استفاده شد نشان داد. به این شکل که شرکت‌های نوپا در هنگام دریافت گواهی دانش‌بنیانی درآمدی از محصولات دانش‌بنیان خود نداشتند، اما شرکت‌های دانش‌بنیان صنعتی که اغلب سابقه بیشتری دارند در هنگام دریافت این گواهی، درحال فروش محصول بودند.



نمودار (1) توزیع تعداد شرکت‌ها در خوشه‌ها به تفکیک زیردسته‌های فناوری

Chart (2) Distribution of companies in clusters by technology subclasses



نمودار (2) توزیع تعداد محصولات در خوشه‌ها به تفکیک زیردسته‌های فناوری

Chart (3) Distribution of products in clusters by technology subclasses

نتیجه‌گیری و پیشنهادها

هدف این مقاله خوشه‌بندی شرکت‌های دانش‌بنیان کشاورزی بر مبنای سه ویژگی زیردستهٔ فناوری، تعداد محصولات دانش‌بنیان شرکت و مقدار فروش این محصولات در شرکت بود. برای

دستیابی به این هدف، از سه روش خوشه‌بندی تفکیکی K-میانگین ساده و K-میانگین در ترکیب با الگوریتم‌های سلسله مراتبی و کلونی مورچه‌ها استفاده شد و بر مبنای شاخص ضریب نیمرخ و توزیع متوازن شرکت‌ها در خوشه‌ها، روش K-میانگین ترکیبی با الگوریتم کلونی مورچه‌ها به عنوان بهترین روش برای این مطالعه انتخاب شد که در آن شرکت‌های دانش‌بنیان کشاورزی ایران به شش خوشه تقسیم‌بندی شدند. در ادامه با توجه به نتایج به دست آمده، پیشنهادهایی ارایه می‌شود.

براساس نتایج به‌دست آمده، لازم است این سیاست‌ها در بخش کشاورزی حداقل برای شش گروه اتخاذ گردد، به طوری که شرکت‌های درون هر خوشه که بیشترین شباهت را با یکدیگر و بیشترین تفاوت را با شرکت‌های خوشه‌های دیگر دارند، از حمایت‌های ویژه خوشه خود برخوردار شوند. برای مثال شرکت‌های خوشهٔ سه که دارای میانگین درآمد پایین‌تری بوده و اغلب جزو شرکت‌های دانش‌بنیان نوپا هستند، اغلب با چالش تقاضا روبه‌رو بوده و نیازمند حمایت‌هایی در جهت معرفی محصول و ورود محصول به بازار هستند و در مقابل، شرکت‌های خوشه‌های چهار و پنج که تنوع محصول و مقدار فروش بیشتری دارند، با ریسک کمتری برای ادامهٔ فعالیت خود روبه‌رو هستند که ضروری است با حمایت‌های حقوقی و بازرگانی به ورود این شرکت‌ها به بازار صادراتی و توسعهٔ بازار کمک شود. از آنجا که شرکت‌های خوشه‌های یک، دو، چهار و پنج بیشترین عضو را در زیردسته‌های یک تا چهار دارند که محصولات آنها اغلب نهاده‌های مصرفی بخش کشاورزی مانند بذر، کود و آفت‌کش‌ها هستند، پیشنهاد می‌شود، شرایطی فراهم شود تا ارتباط این شرکت‌ها با بازار هدف که کشاورزان هستند تسهیل گردد. برای تحقق بخشیدن این امر، برگزاری نمایشگاه‌هایی برای معرفی محصولات در مناطق مختلف کشور ضرورت دارد. برای اعضای خوشهٔ شش که زیردستهٔ فناوری محصولات آن اغلب دورهٔ آزمایشگاهی پیش از تولید طولانی‌تری دارند، تقویت حمایت‌های فنی از جمله تسهیلات آزمایشگاهی و ایجاد ارتباط با سرمایه‌گذاران برای پروژه‌های تحقیق و توسعه محصولات این شرکت‌ها دارای اهمیت فراوانی است.

با توجه به آنچه بیان شد، ارایه تسهیلات به شرکت‌های دانش‌بنیان کشاورزی نیز نمی‌تواند، تابع الگوی یکسانی برای همه باشد و ضروری است، سطح و میزان ارایه تسهیلات و نحوهٔ بازپرداخت آن‌ها با توجه به نتایج خوشه‌بندی، انجام پذیرد و مطالعات تکمیلی در این زمینه برنامه‌ریزی شود.

منابع

Alfansi, L. and Sargeant, A, (2000). Market segmentation in Indonesian banking sector: The relationship between demographics and desired customer benefits. International Journal of Bank Marketing, 18(2): 86-98.

Aliheidari Bioki، T. and Khademizare, H. (2015). Improvement of DEA approach for clustering credit rating of customer in banks, journal of Modelling in Engineering, 13(41): 59-74.

Brandt, A., (2005). Cluster analysis for market segmentation. Department of Curriculum Teaching and Learning, Toronto, University of Toronto. Master of Art.

Chen, J., Z. Zhu and H. Y. Xie (2004). Measuring intellectual capital: a new model and empirical study, Journal of Intellectual Capital, 5(1): 195-212.

Daie, P., Li, S. (2016). Hierarchical clustering for structuring supply chain network in case of product variety, Journal of Manufacturing Systems, 38: 77-86.

Epetimehin, F.M., (2011). Market segmentation: A tool for improving customer satisfaction and retention in insurance service delivery .Journal of Emerging Trends in Economics and Management Sciences (JETEMS), 2(1): 62-67.

Golichenar, F. and Saniee Abadeh, M. (2012). An overview of the combined methods of k-means clustering algorithms and meta-revelation algorithms to confront the disadvantages of the k-means algorithm, Regional Conference on Computer Science, Computer Engineering and Information Technology, Fars, Iran.

Kuo, R. J., H. S. Wang, Tung-Lai Hu, S. H. Chou. (2005). Application of ant K-means on clustering analysis, Computers & Mathematics with Applications, 50 (10–12): 1709-1724.

Linares-Mustarós, S., Coenders, G., Vives-Mestres, M. (2018). "Financial performance and distress profiles. From classification according to financial ratios to compositional classification". Advances in Accounting, DOI: 10.1016/j.adiac.2017.10.003

Momeni, M. (2014). Data clustering, Moalef Publication, Tehran, Iran(In Farsi).

OECD. (2007). Competitive Regional Clusters. Paris: OECD.

Rahman Md., K.Z. Arefin · S. Masud · S. Sultana · R.M. Rahman. (2017). Analyzing Life Insurance Data with Different Classification

Techniques for Customers' Behavior Analysis, Advanced Topics in Intelligent Information and Database Systems, Studies in Computational Intelligence 710, DOI 10.1007/978-3-319-56660-3_2.

Sadeghian, M. (2013). Finding the optimal center for clusters in the K-means algorithm, MSc. Thesis, University of Sheikh-e-Bahaeei (In Farsi).

Salmani, B. (2015). Science and Technology Parks Mechanism for Knowledge Economy, Nooreelm Publication, Tehran, Iran (In Farsi).

Shirkavand, S., Jokar, GH. and Mazidi, A. (2014). Market Segmentation of Insurance Customers Based on Expected Values, Iranian Journal of Insurance Research, 29 (116): 137-163 (In Farsi).

Vice Presidency for Science and Technology of Islamic Republic of Iran. (2017). Statistics of Knowledge-based Companies in Tehran (In Farsi).

Vice Presidency for Science and Technology of Islamic Republic of Iran. (2017). Table of Knowledge-based Goods and Services, forth edited, Tehran, Iran (In Farsi).

Zare AhmadAbadi, H., Emam, M. and Naser Sadrabadi, A. (2017). Market Clustering with Ant Colony Optimization (Comparative approach with k-means), JOURNAL OF BUSINESS ADMINISTRATION RESEARCH, 8 (16): 17-36 (In Farsi).

DOI :10.22034/IAES.2019.101251.1665

Agricultural Economics/Volume 13/Issue 2/1398

Application of Ant Colony and Hierarchical Metaheuristic Algorithms in segmentation of Agricultural Knowledge Based Companies

Elham Mehrparvar Hosseini, Hamed Rafiee, Seyed Safdar Hosseini, Amirhossein Chizari[1]

Received: 06 Jan.2019 **Accepted:15 July.2019**

Extended Abstract

Introduction

Clustering is one of the most important operations in data mining and its results are useful for researchers and policy makers in various fields for analysis and planning. Since in recent years, the knowledge based economy has been developing with the support of knowledge-based companies in Iran, the analysis of the characteristics of these companies and their segmentation for effective planning provides an appropriate perspective for policy makers.

Materials and Methods

In this article, Iranian knowledge based companies of agriculture using 2017 data have been clustered based on the field of technology, the number of products and value of product sales. Three clustering methods, simple K-Means and K-Means in combination with hierarchical and ant colony algorithms by using MATLAB 2016a software were applied for clustering. Then, results of three algorithms have been compared and the best one for this data have been selected.

Results and discussion

According to the results, the clustering using K-Means in combination with the ant colony algorithm, in comparison with the other two methods, shows a more balanced distribution of the firms among six clusters, and the average Silhouette width value of the 0.7 confirms the validity of this clustering. The highest number of companies and products are in first

[1]Respectively: Ph.D. Candidate Assistant Professor,Professor,Assistant Professor of Agricultural economics,University of Tehra
Email: hamedrafiee@ut.ac.ir

cluster and the lowest are in 4th and 5th clusters which have the highest average sales per company with 27293 and 5404 million Rials, respectively. The reason for the small number of members in these clusters is that few companies have a large number of knowledge-based products with high sale. In contrast, most companies have acquired a small market share with few number of products. The variety of products and sales of companies indicates the flexibility of the company in different markets, the size of the market and the ability to participate in market development.

Suggestion

Based on the results, it is suggested that similar support programs be designed for companies that are due to similar characteristics in a cluster in order to be effective and avoid losing budget. For knowledge based companies with low average sales that are in a cluster, market support programs and for companies with products diversification and high average incomes, export market development programs are recommended. Providing banking facilities to agricultural knowledge-based companies can not have the same model for every company, and It is necessary to determine the amount of facilities provided and the method of repayment based on the results of the clustering..

JEL Classification: C81, L2, M13, O31, Q16
Keywords: Knowledge-based Economics, Data Mining, K-Means Clustering, Hierarchical Algorithm, Ants colony Algorithm