

خوشه‌بندی فروشگاه‌های آنلاین از نگاه تأمین‌کننده با کمک بهینه‌یابی تعداد خوشه‌ها در الگوریتم دو مرحله‌ای SOM

محمدعلی فائزی راد،* علیرضا پویا**

تاریخ دریافت: ۹۴/۱۱/۱۰

تاریخ پذیرش: ۹۵/۴/۲۱

چکیده

با گسترش فناوری اطلاعات و ظهور بازارهای مجازی، برنامه‌ریزی و تحلیل این بازارها و اجزای آنها به اولویت سازمان‌های ذی‌نفع تبدیل شده است. یکی از مهم‌ترین بازارهای مجازی کنونی در کشور، بازار فروش آنلاین خدمات شارژ اپراتورهای مختلف تلفن همراه در قالب فروشگاه‌های اینترنتی است. با توجه به تعداد انبوه و در حال رشد این فروشگاه‌ها، دسته‌بندی آنها از دید تأمین‌کننده به منظور ارائه امتیازات و آینده‌نگری در همکاری ضروری است. در این پژوهش، با استفاده از داده‌های یکی از مهم‌ترین تأمین‌کنندگان این محصولات، قریب به سه هزار فروشگاه مجازی مورد تحلیل قرار گرفته و بر اساس شاخص‌های مورد نظر تأمین‌کننده خوشه‌بندی شده است. فرآیند خوشه‌بندی با استفاده از شبکه عصبی مصنوعی SOM و در قالب روش دو مرحله‌ای آن با الگوریتم *k-means* انجام شده است که تحلیل خوشه‌های حاصل از شبکه عصبی را تسهیل می‌کند. اگرچه به منظور تعیین بهترین تعداد خوشه‌ها، شاخص‌های اعتبارسنجی مختلفی گسترش پیدا کرده است اما در این پژوهش با ترکیب روش تصمیم‌گیری چندشاخصه در مدل و تجمیع شاخص‌های مختلف، به ارائه یک مدل بهینه‌یابی با رویکرد جبرانی نسبت به شاخص‌ها پرداخته شده است.

واژگان کلیدی: تحلیل خوشه‌ای، تصمیم‌گیری چندشاخصه، شبکه عصبی مصنوعی، فروشگاه آنلاین، نگاهت خود سازمان ده (SOM)

* دانشجوی دکتری مدیریت تحقیق در عملیات، دانشکده علوم اداری و اقتصاد، دانشگاه فردوسی مشهد، ایران

** دانشیار گروه مدیریت، دانشکده علوم اداری و اقتصاد، دانشگاه فردوسی مشهد، ایران (نویسنده مسئول)

مقدمه

امروزه بازارهای مصرفی با تغییرات فناوری، محدودیت‌های زمانی، تنوع محصولات و عوامل متعدد دیگر، اشکال متنوع و گوناگونی از خرده‌فروشی نظیر سیستم‌های سنتی، کاتالوگی، سیستم‌های آنلاین و ... به خود می‌بیند (شیم، استلیک، لاتز و وارینگتون، ۲۰۰۱). سیستم‌های آنلاین در بستر اینترنت به جایگاهی گسترده برای معاملات تجاری تبدیل شده است و کسب و کارها به سمت الکترونیکی شدن پیش می‌روند و یکی از جنبه‌های مهم این موضوع پیدایش فروشگاه‌های اینترنتی و مجازی است (منتظری، ابراهیمی، احمدی و راهنما، ۱۳۹۳). به خصوص زمانی که محصول مورد نظر کاربر اینترنتی، خود یک کالای مجازی یا شبه‌مجازی باشد.

در دهه اخیر، با گسترش پرشتاب تلفن همراه در کشورمان و به تبع آن تنوع اپراتورهای تلفن همراه، بازار بسیار گسترده‌ای برای فروش شارژ سیم‌کارت‌های اعتباری این اپراتورها و محصولات جانبی ایجاد شده است. این بازار اگرچه امروزه در قالب‌های مختلفی نظیر فروش مستقیم، فروش از طریق پایانه‌های فروش (POS)، از طریق خودپردازها و یا کدهای USSD جریان دارد، اما همچنان یکی از گسترده‌ترین بخش‌های تأمین این نیاز خریداران از طریق وب‌سایت فروش این محصولات در فضای اینترنت تأمین می‌گردد. این وب‌سایت‌ها، اغلب به یکی از تأمین‌کنندگان اصلی متصل هستند و عملاً واسطه فروش محصول آن تأمین‌کننده در سایت خویش محسوب می‌شوند.

در این پژوهش، با استفاده از پایگاه داده یکی از تأمین‌کنندگان مطرح این بازار، تحلیل بازار خرده‌فروشان مجازی متصل به آن صورت گرفته است. بنابراین، نگاه تحلیل از منظر ادراکات و ارزش‌های تأمین‌کننده است و کاربران آن - که خود با صدها و هزارها مشتری سر و کار دارند - همان فروشگاه‌های مجازی هستند. تحلیل بازار در این پژوهش، با استفاده از رویکرد تحلیل خوشه‌ای صورت گرفته است. اهمیت این خوشه‌بندی برای تأمین‌کننده، تفکیک مشتریان به منظور ارائه تسهیلات مالی و خدماتی به آنان است. با تحقق گروه‌بندی مشتریان،

فرآیند توسعه فروشگاه‌ها جهت‌دار شده و تعیین استراتژی رقابتی با رویکردهای تشویقی هدف‌دار خواهد شد.

در این پژوهش با استخراج شاخص‌ها توسط تأمین‌کننده، معیارهای اصلی برای خوشه‌بندی فروشگاه‌ها تعیین شده و سپس با استفاده از رویکرد نوینی اقدام به دسته‌بندی فروشگاه‌ها می‌شود. این رویکرد بر اساس ادغام دو مرحله‌ای شبکه عصبی نگاشت خودسازمان‌ده و روش K-means است. این ترکیب موجب می‌شود که تحلیل خوشه‌های شبکه عصبی تسهیل گردد و بدون نیاز به تفسیر شهودی با رویکردی متقن به تجمیع خوشه‌ها پرداخته شود. در بسیاری از پژوهش‌های مبتنی بر خوشه‌بندی با شبکه عصبی، خوشه‌های حاصل صرفاً بر اساس نظر خبرگان تجمیع شده که عاری از خطا نیست. مزیت دیگر این پژوهش ارائه مدلی نوین به منظور تعیین تعداد خوشه‌ها در بازه قابل قبول خبرگان با در نظر گرفتن شاخص‌های متنوع اعتبارسنجی خوشه‌ها و بهینه‌یابی ساختار قابل قبول با تکنیک تصمیم‌گیری چندشاخصه است. در بسیاری از پژوهش‌ها، انتخاب بهترین تعداد خوشه‌ها صرفاً بر اساس یک شاخص است و یا اینکه از چندین شاخص به صورت همزمان اما با رویکرد غیرجبرانی استفاده می‌شود که این موضوع در پژوهش حاضر، مرتفع شده و بهره‌گیری از رویکرد MADM تجمیع شاخص‌ها را ممکن می‌سازد.

مبانی نظری

بخش‌بندی مشتریان

تحلیل رفتار مشتری یکی از راه‌های شناخت هرچه بیشتر بازار و ظرفیت‌های آن است. از همین روی، استفاده از داده‌های اساسی برای ارتباط بیشتر با مشتریان و ویژگی‌های آنان، ابزار مهمی در مدیریت ارتباط با مشتری است. بنابراین، استفاده از پایگاه‌های اطلاعاتی که دربردارنده داده‌های مربوط به مشتریان است، می‌تواند در حکم سرمایه‌ای قابل توجه باشد (خدیور، رزمی و حامدی، ۱۳۹۲). بخش‌بندی مشتریان به مفهوم تقسیم بازار هدف به زیرمجموعه‌هایی است

که هر یک از آنها را می‌توان با استراتژی متناسب راهبری کرد (کاتلر و آرمسترانگ، ۱۹۹۹). بخش‌بندی مشتریان یک گام اساسی در تبدیل داده‌های انبوه و پیچیده به داده‌های قابل برنامه‌ریزی خواهد بود. با دسته‌بندی بر اساس شباهت‌های مشتریان می‌توان هزینه‌ها را کاهش داده و رضایت آنان را بیشتر تأمین نمود (وانگ، ما، لائو و وانگ، ۲۰۱۴). تفکیک مشتریان با رویکرد داده‌کاوی موجب خواهد شد که افزایش ارزش مشتری در حوزه مربوطه و در سطح معین آن مد نظر قرار گیرد (لیانگ، ۲۰۱۰). شناخت درست مشتریان این امکان را ایجاد می‌کند که شرکت‌ها نسبت به ارائه تعرفه خاص، امتیاز و امکاناتی مطابق با ویژگی‌های مشتریان تفکیک شده غیرمعمولی و خاص اقدام کنند (لوپز و همکاران، ۲۰۱۱).

نگاشت خودسازمان‌ده دو مرحله‌ای

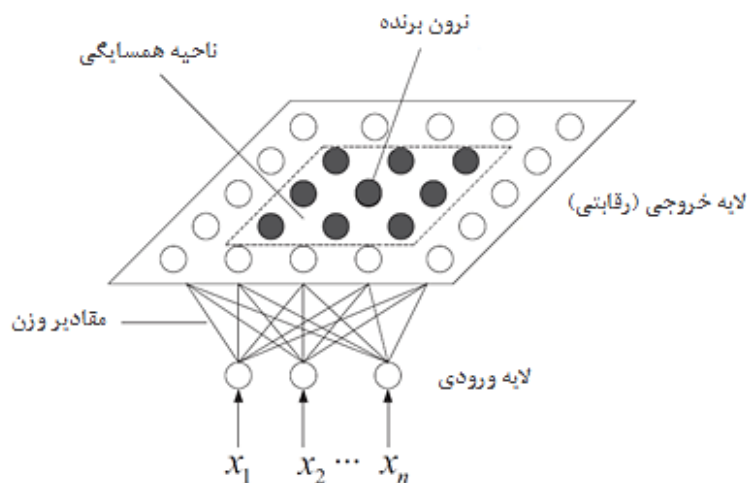
تحلیل خوشه‌ای یا خوشه‌بندی یکی از اساسی‌ترین شاخه‌های یادگیری بدون نظارت در داده‌کاوی است. تحلیل خوشه‌ای در حوزه‌های مختلفی کاربرد که دارد که یکی از مهم‌ترین آنها تحقیقات بازار و مطالعات بازاریابی است (شاه‌بابا و بهشتی، ۲۰۱۴). خوشه‌بندی، نوعی عملیات داده‌کاوی غیرمستقیم است. در اکثر روش‌های داده‌کاوی با یک مجموعه آموزشی آغاز به کار می‌شود که به کمک آن سعی می‌کنیم مدلی را ایجاد نماییم تا داده‌ها را بخش‌بندی کرده و سپس داده‌های جدید را پیش‌بینی کنیم؛ اما در خوشه‌بندی هیچ دسته‌ای از قبل وجود ندارد و در واقع متغیرها به دو طبقه مستقل و وابسته تقسیم نمی‌شوند. در خوشه‌بندی تمرکز بر روی گروه‌هایی از داده‌هاست که به شیبه هستند تا با کشف این شباهت‌ها بتوان رفتارها را بهتر شناسایی کرده و بر مبنای این شناخت به تصمیم بهتری رسید (غضنفری، علیزاده و تیمورپور، ۱۳۸۷). روش‌های مختلفی با رویکردهای متفاوت برای انجام تحلیل خوشه‌ای داده‌ها توسعه داده شده‌اند که از جمله آن‌ها رویکرد نگاشت خودسازمان‌ده با محوریت شبکه عصبی مصنوعی است.

شبکه نگاشت خودسازمان‌ده یا به اختصار SOM می‌تواند فرآیند یادگیری را بر روی داده‌های پیچیده و چندبُعدی انجام دهد و در نتیجه مجموعه‌ای قابل مشاهده از خوشه‌ها استخراج کند (کوهونن، ۱۹۸۹).

آموزش شبکه SOM از دو فاز تکرارشونده تشکیل شده است: فاز نخست عبارت است از انتخاب بهترین واحد نگاشت (بهترین نرون شبکه عصبی) به منظور انطباق با هر یک از داده‌های ورودی، و فاز دوم به روزرسانی این نگاشت به منظور ارائه بهترین نمایندگی و نمایش از داده‌های ورودی است (کوهونن، ۱۹۹۸). فرآیند انتخاب بهترین واحد به منظور انطباق با داده‌های ورودی (بهترین واحد انطباقی یا BMU^1) بر اساس کمترین مقدار فاصله - معمولاً فاصله اقلیدسی - انجام می‌شود. سپس در فاز به‌روزرسانی، هر BMU و واحدهای همسایه آن (در شعاع معین) به نزدیکی داده ورودی و انطباق کامل با آن حرکت می‌کنند. این شعاع همسایگی، با هر فاز انتخاب و به‌روزرسانی کمتر می‌شود تا در نهایت به یک نگاشت نهایی - دو بُعدی - منجر گردد (کلارک، سارلین، شارما و سیسون، ۲۰۱۵).

فرآیند آموزش شبکه SOM در قالب گام‌های زیر قابل انجام است (یوسف، ۲۰۰۶).
گام ۱). تعداد نرون‌های لایه ورودی (داده‌های ورودی) و لایه رقابتی (نگاشت نهایی) به ترتیب برابر با n و m فرض می‌شود و سپس توپولوژی دو بُعدی شبکه مطابق شکل ۱ ایجاد می‌شود. تمامی وزن‌های اولیه که بر اتصال لایه ورودی و لایه رقابتی اعمال می‌شوند، اعداد تصادفی در بازه ۰ و ۱ هستند.

^۱. Best Matching Unit (BMU)



شکل ۱. ساختار شبکه SOM (وانگ، بیان، لیو و لیو، ۲۰۱۳)

گام ۲). به منظور محاسبه فاصله بین بردارهای ورودی و بردارهای وزن از فاصله اقلیدسی مطابق رابطه ۱ استفاده می‌شود که در آن $X_k = [x_{k1}, x_{k2}, \dots, x_{kn}]$ و $W_j = [x_{j1}, x_{j2}, \dots, x_{jn}]$ به ترتیب نماینده بردار ورودی k ام ($0 \leq k \leq K$) و بردار وزن آن است.

$$d_{jk} = \|X_k - W_j\| = \sqrt{\sum_{i=1}^n (X_{ki} - W_{ji})^2} \quad j = 1, 2, 3, \dots, m \quad (1)$$

بر همین اساس، نرون برنده یا به عبارت دیگر BMU که کم‌ترین فاصله را تا بردار X_k دارد بر اساس رابطه ۲ محاسبه می‌شود.

$$\|X_k - W_c\| = \min\{d_{jk}\} \quad (2)$$

این رابطه مبین انتخاب BMU است که در آن نرون برنده با C مشخص شده و فاصله آن تا بردار k ام، حداقل فاصله است.

گام ۳. به روزرسانی وزن نرون برنده و همسایگان آن، گام بعدی الگوریتم است. پس از مشخص شدن نرون‌های برنده، SOM بردارهای وزن را برای نرون برنده به‌روز می‌کند. بردارهای وزن در جهت بردار ورودی تقویت شده و با اعمال رابطه ۳ به‌روز می‌شوند.

$$W_j(t+1) = W_j(t) + \eta(t)h_{jc}(t)[X_k(t) - W_j(t)] \quad (3)$$

در این رابطه، $W_j(t)$ و $W_j(t+1)$ نماینده بردار وزن نرون خروجی j ام به ترتیب در قبل و بعد از به‌روزرسانی وزن است. همچنین، $\eta(t)$ نرخ یادگیری در بازه ۰ و ۱ است و $h_{jc}(t)$ نمایش‌دهنده تابع همسایگی است.

برای تعیین تابع همسایگی می‌تواند از توابع مختلفی استفاده نمود که از جمله مشهورترین آنها، تابع گوسین به شرح رابطه ۴ است.

$$h_{jc}(t) = \exp\left(-\frac{\|r_c - r_j\|}{\sqrt{2}\sigma(t)}\right) \quad (4)$$

که در آن، r_c و r_j به ترتیب نشان‌دهنده مکان قرارگیری نرون برنده و سایر نرون‌ها در نقشه دوبعدی با شعاع همسایگی σ است.

گام ۲) و ۳) برای تمامی بردارهای ورودی انجام می‌شود تا تمامی نرون‌های لایه خروجی به‌طور پیوسته بر نمونه‌های لایه ورودی متناظر گردند.

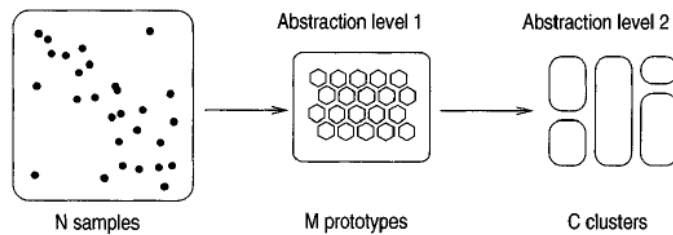
گام ۴). به منظور اجرای مرحله بعد، لازم است که نرخ یادگیری و تابع همسایگی به‌روز گردد.

$$\eta(t) = \eta_0 \left(1 - \frac{t}{T}\right) \quad (5)$$

$$\begin{cases} \sigma(t) = \sigma_0 \exp(-t/\tau) \\ h_{jc}(t) = \exp\left(-d_{ij}^2 / \sqrt{2}\sigma(t)\right) \end{cases} \quad (6)$$

در رابطه فوق، σ مقدار اولیه σ را نشان می‌دهد و τ ثابت زمانی است. با افزایش گام‌های یادگیری، شعاع همسایگی کاهش می‌یابد تا در نهایت به صفر برسد. مراحل فوق تا رسیدن به تکرار T ام تکرار می‌شود تا در نهایت به یک الگوی معین برسد.

روش SOM دو مرحله‌ای یک روش خوشه‌بندی است که بر اساس ترکیب دو روش SOM و K-means به وجود آمده است. این روش، در ابتدا با استفاده از الگوریتم SOM اقدام به خوشه‌بندی داده‌ها می‌کند، سپس با روش K-means مراکز اولیه (نرون‌های شبکه عصبی) در توپولوژی SOM را خوشه‌بندی می‌کند (وسانتو و الهونیمی، ۲۰۰۰). مراحل این روش، در شکل ۲ به تصویر کشیده شده است.



شکل ۲. روش دو مرحله‌ای نگاشت خودسازمان‌ده (وسانتو و الهونیمی، ۲۰۰۰)

الگوریتم k-means یکی از معروف‌ترین و محبوب‌ترین روش‌های خوشه‌بندی است که از k مرکز اولیه شروع می‌کند و با روندی تکرارشونده به بهینه‌سازی کیفیت خوشه‌بندی می‌پردازد (کالیانی و اسووراپ، ۲۰۱۱). این الگوریتم تلاش می‌کند که مراکز خوشه‌ها (C_1, C_2, \dots, C_K) را در شرایطی که مجموع مربع فواصل هریک از نقاط داده (x_i) از نزدیک‌ترین مرکز خوشه (C_k) کمترین مقدار را داشته باشد (ردموند و هنگان، ۲۰۰۷). این فرآیند در رابطه ۷ فرموله شده است.

$$D = \sum_{i=1}^n \left[\min_{k=1,2,\dots,K} d(x_i, c_k) \right]^2 \quad (7)$$

شاخص‌های اعتبارسنجی

بسیاری از روش‌های خوشه‌بندی به تعداد از پیش تعریف‌شده برای خوشه‌ها نیاز دارند؛ اما در دنیای واقعی، تعداد درست خوشه‌ها (CNC)^۱ یک مقدار از قبل تعیین‌شده نیست. از این روی، چالش اصلی این روش‌های خوشه‌بندی تعیین این تعداد است (شاه‌بابا و بهشتی، ۲۰۱۴). روش K-means نیز از جمله این روش‌های خوشه‌بندی است که رویکردهای متعددی به منظور ارزیابی تعداد پیشنهادی خوشه‌ها برای آن گسترش یافته است (چیانگ و میرکین، ۲۰۱۰). از جمله روش‌های پیشگام در این حوزه می‌توان به شاخص‌های اعتبارسنجی اشاره کرد. هر شاخص اعتبارسنجی قادر است که برای هر CNC مقدار متناظرش را محاسبه کند و سپس بر اساس انتخاب مقدار بهینه خود، بهترین تعداد خوشه را معرفی کند (شاه‌بابا و بهشتی، ۲۰۱۴).

شاخص‌هایی که در این پژوهش مورد استفاده قرار گرفتند، در ادامه به صورت مجزا معرفی می‌شوند.

شاخص دان^۲

شاخص دان (دان، ۱۹۷۴) با هدف تعیین فشردگی و مجزا بودن خوشه‌ها ارائه شده است. شیوه محاسبه شاخص دان بر اساس

$$DI = \min_{1 \leq i \leq K} \left\{ \min_{i+1 \leq j \leq K} \left\{ \frac{d(C_i, C_j)}{\max_{1 \leq l \leq K} diam(C_l)} \right\} \right\} \quad (۸)$$

1- correct number of clusters

2- Dunn index

است که در آن $d(C_i, C_j)$ نشان‌دهنده مقدار فاصله بین خوشه i و j بوده و $diam(C_l)$ نماینده قطر خوشه l است. این دو مقدار به صورت زیر محاسبه می‌شوند:

$$d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y) \quad (9)$$

$$diam(C_l) = \max_{x, y \in C_l} d(x, y) \quad (10)$$

هرچه مقدار این شاخص بیشتر باشد، نشان‌دهنده فشردگی بیشتر اعضای خوشه‌ها و مجزاتر بودن خود خوشه‌ها است؛ لذا مقدار بهینه این شاخص، بیشینه آن است.

شاخص نیمرخ^۱

شاخص نیمرخ بر اساس ماتریس عدم تشابه بنا شده است (مؤمنی، ۱۳۹۰). مقدار نیمرخ برای هر نقطه بر اساس میزان شباهت آن به خوشه‌اش نسبت به میزان شباهت به خوشه‌های دیگر محاسبه می‌شود.

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (11)$$

در این رابطه، a_i نمایش‌دهنده میانگین فاصله از نقطه i ام تا سایر نقاط هم‌خوشه‌ای آن است و b_i نماد کمترین مقدار متوسط فاصله میان نقطه i ام تا سایر خوشه‌هاست (روسیوف، ۱۹۸۷). شاخص نیمرخ در بازه $+1$ تا -1 قرار می‌گیرد که بزرگترین مقدار آن، بهترین تعداد خوشه‌ها را نشان می‌دهد.

شاخص دیویس - بولدین (DB)^۲

شاخص DB بر اساس نسبت فواصل درون‌خوشه‌ای و برون‌خوشه‌ای ایجاد شده است. این نسبت با استفاده از

1- Silhouette

2- Davies-Bouldin

$$DB = \frac{1}{k} \sum_{i=1}^K \max_{i \neq j} \{D_{i,j}\} \quad (12)$$

محاسبه می‌شود که در آن نسبت فاصله درون‌خوشه‌ای به برون‌خوشه‌ای است (دیویس و بولدین، ۱۹۷۹). هرچه مقدار این شاخص بیشتر شود، از کیفیت خوشه‌بندی کاسته شده و لذا بهترین خوشه‌بندی دارای کمترین مقدار شاخص DB است.

شاخص کالینسکی - هاراباسز (VRC) ۱

این شاخص علاوه بر عنوان فوق، گاهی معیار نسبت واریانس (VRC) نیز نامیده می‌شود. شیوه محاسبه این شاخص با استفاده از

$$VRC_k = \frac{SS_B}{SS_W} \times \frac{(N - k)}{(k - 1)} \quad (13)$$

است که در آن SS_B مجموع واریانس بین خوشه‌ها و SS_W مجموع واریانس درون خوشه‌هاست. N و k نیز به ترتیب تعداد مشاهدات و تعداد خوشه‌ها را نشان می‌دهند (کالینسکی و هاراباسز، ۱۹۷۴). بهترین تعداد خوشه‌ها بر اساس بزرگترین مقدار VRC تعیین می‌گردد.

شاخص RS ۳

این شاخص، نسبت مجموع توان دوم انحرافات بین گروه‌ها (SS_B) به مجموع توان دوم انحرافات کل داده‌ها (SS_T) است. به این شاخص ضریب تشخیص نیز گفته می‌شود.

$$RS = \frac{SS_B}{SS_T} = \frac{SS_T - SS_W}{SS_T} \quad (14)$$

1- Calinski-Harabasz
2- variance ratio criterion
3- R-Squared

دامنه تغییرات این ضریب بین ۰ و ۱ است و هرچه به یک نزدیک تر باشد بیانگر آن است که پراکندگی داده‌های داخل یک خوشه کم و فاصله بین خوشه‌ها زیاد است (مؤمنی، ۱۳۹۰). بنابراین، بهترین ساختار خوشه‌بندی متناظر با بزرگترین مقدار RS خواهد بود.

پیشینه پژوهش

بهره‌گیری از رویکرد نگاشت خودسازمان‌ده (SOM) در پژوهش‌های مختلف به منظور خوشه‌بندی بازار و مشتریان رایج است. از جمله این پژوهش‌ها می‌توان به بهره‌گیری از این روش در حوزه فناوری اطلاعات اشاره نمود.

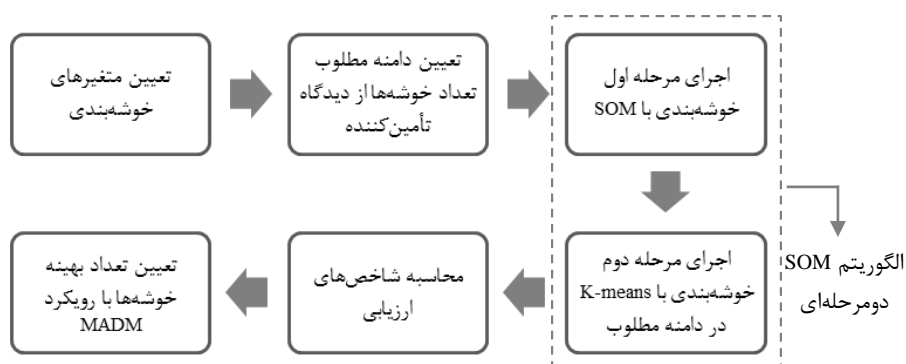
تقوا و حسینی بامکان (۱۳۹۰) رفتار مشتریان را در سیستم بانکداری الکترونیک با استفاده از رویکرد SOM تحلیل کرده‌اند. در این پژوهش، متغیرهای مختلفی برای هریک از مشتریان بانکداری الکترونیک یک بانک تعریف شده و بر مبنای اقدام به خوشه‌بندی این مشتریان شده است. افجه و درویشی (۱۳۹۳) در پژوهش دیگری، اقدام به بخش‌بندی بازار در یک محیط چندکاناله بر اساس ویژگی‌های مصرف‌کنندگان کرده‌اند. بر اساس آنچه در مقاله فوق آمده است، پژوهشگران بر مبنای نگرش به کانال خرده‌فروشی برای جستجوی اطلاعات و خرید از کانال‌های اینترنت، کاتالوگ و فروشگاه، مصرف‌کنندگان را بخش‌بندی کرده‌اند. در پژوهش دیگری، تاونماس، کوراشیگ، لیزوکا و کانتارزیک (۲۰۰۶) با استفاده از شبکه عصبی SOM اقدام به تحلیل خوشه‌ای کاربران بازی‌های آنلاین کرده‌اند که در آن ورودی این شبکه را میزان دنبال کردن در نقشه بازی توسط کاربران قرار داده‌اند. هانگ و تسای (۲۰۰۸) طی پژوهشی، عوامل مؤثر بر بخش‌بندی بازار را بر اساس تقاضای بازار رسانه تحلیل کرده و با استفاده از ارائه تحلیلی جدید بر مبنای SOM سلسله‌مراتبی، تحلیل خوشه‌ای این بازار را ارائه کردند. از سوی دیگر، کیم و آن در سال ۲۰۰۸ در مقاله‌ای مدل نوینی را بر مبنای روش خوشه‌بندی K-means، الگوریتم ژنتیک و شبکه SOM به منظور بخش‌بندی بازار تجارت الکترونیک و فروش آنلاین ارائه کردند.

روش‌شناسی پژوهش

پژوهش حاضر به دنبال خوشه‌بندی فروشگاه‌های آنلاین و یا کاربران خرده‌فروش یک تأمین‌کننده است. داده‌های این پژوهش مربوط به ۳۰۱۱ فروشگاه آنلاین کارت شارژ الکترونیکی و محصولات وابسته آن است که توسط یکی از شرکت‌های مطرح تأمین‌کننده در این بازار تأمین می‌شوند. این فروشگاه‌ها در حقیقت بخش B2B کسب و کار این شرکت را تشکیل داده و هر یک با راه‌اندازی کسب و کار خویش در طول تأمین کالا توسط شرکت، تبادل مالی متقابل را کلید زده‌اند. با توجه به دربرگیری همه فروشگاه‌ها، نمونه‌گیری بر پایه سرشماری کامل صورت پذیرفته است. پس از بررسی پیشینه پژوهش و تایید نظرات کارشناسان تأمین‌کننده، این عوامل به عنوان شاخص‌های خوشه‌بندی فروشگاه‌ها مد نظر قرار گرفته است: تاریخ اولین ورود (عمر فروشگاه)، تاریخ آخرین ورود (آخرین فعالیت)، تاریخ آخرین فروش، مدت فعالیت مفید، مبلغ فروش یک ماه اخیر، مبلغ رشد نسبت به ماه قبل، مبلغ فروش یک هفته اخیر، مبلغ رشد نسبت به هفته قبل. بر این اساس، اطلاعات مربوط به فروشگاه‌های فوق از پایگاه داده‌های شرکت تأمین‌کننده استخراج شده و مورد تحلیل قرار گرفت. لازم به ذکر است که تاریخ‌ها بر حسب تعداد روز تا روز مرجع (زمان پژوهش) محاسبه و در پایگاه داده‌های پژوهش ثبت شد.

تحلیل خوشه‌ای پس از تعیین ویژگی‌ها و مشخصه‌های شباهت‌ها و تفاوت‌ها میان این فروشگاه‌ها توسط تأمین‌کننده صورت می‌گیرد. بعد از تعیین دامنه تعداد خوشه‌ها توسط تأمین‌کننده، خوشه‌بندی برای هر یک از این تعداد با روش دو مرحله‌ای SOM انجام گرفته و سپس بهترین شاخص‌های ارزیابی خوشه‌بندی برای هر کدام محاسبه می‌شود. این شاخص‌ها بر اساس شیوه محاسبه و منطق خود، هر یک می‌توانند پاسخ متضاد یا متفاوتی با دیگری داشته باشند، از این رو در این پژوهش با رویکردی جدید اقدام به تحلیل شاخص‌ها با استفاده از تکنیک‌های MADM (تصمیم‌گیری چندشاخصه) شده است. با این اقدام، شاخص‌های اعتبارسنجی خوشه‌بندی به عنوان معیارهای MADM و هر یک از تعداد خوشه‌ها به عنوان

گزینه تصمیم‌گیری در نظر گرفته می‌شوند. پس از این کار قادر خواهیم بود که بهترین تعداد خوشه‌بندی را در میان ساختارهای مختلف خوشه‌بندی بیابیم. مراحل کلی این پژوهش در شکل ۳ مشخص شده است.



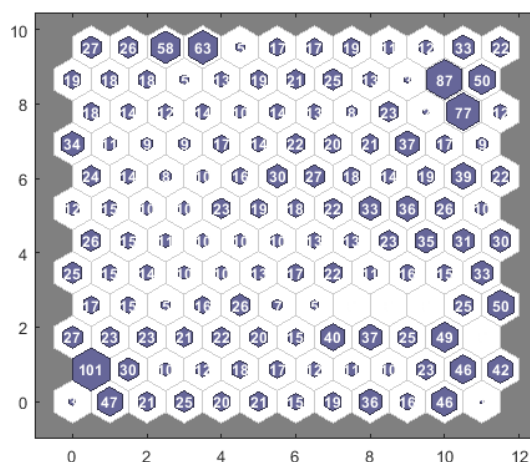
شکل ۳. روند پژوهش به منظور تعیین بهترین ساختار خوشه‌بندی

ابعاد شبکه SOM با توجه به تعداد خوشه‌های اولیه تعیین می‌شود. با توجه به ادبیات پژوهش مبنی بر واگذاری انتخاب تعداد خوشه‌ها به خبرگان و با نظر به اینکه تمایل کارشناسان شرکت - با توجه به انواع تسهیلات قابل ارائه و استراتژی کنونی - درباره تعداد دسته‌های مورد تحلیل، بازه‌ای از تعداد ۵ تا ۱۲ گروه است، لذا در این پژوهش در نهایت باید به تعداد خوشه‌های بهینه در این بازه رسید و عملاً یک انتخاب مقید در این بازه صورت می‌گیرد. با توجه به تکنیک دو مرحله‌ای SOM در گام نخست باید تعداد بیشتری از خوشه‌ها را ایجاد و سپس مراکز خوشه‌های ایجاد شده را با روش K-means مجدداً در بازه ۵ تا ۱۲ خوشه، دسته‌بندی نمود. در نهایت بهترین تعداد خوشه‌ها نیز با استفاده از رویکرد تصمیم‌گیری چندشاخصه انجام می‌پذیرد. در این پژوهش، تکنیک TOPSIS به عنوان روش تصمیم‌گیری چندشاخصه مورد استفاده قرار داده شده است. مبنای این روش، یافتن نزدیک‌ترین گزینه به

گزینه ایده آل مثبت در عین حداکثر فاصله از گزینه ایده آل منفی است (لین، لی، چانگ و تینگ، ۲۰۰۸).

تجزیه و تحلیل و یافته‌های پژوهش

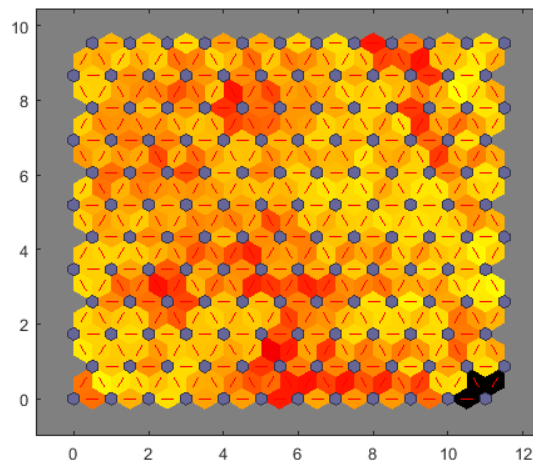
بنا بر توصیه وسانتو و الهونیمی (۲۰۰۰) که تعداد خوشه‌های نهایی را ما بین ۲ تا \sqrt{n} کل داده‌ها می‌داند، می‌توان تعداد اولیه خوشه‌های شبکه SOM (نورون‌ها) را نسبت به همین اعداد و با توجه به بازه مورد نظر خبرگان که در بخش قبل ذکر شد، محاسبه نمود. با توجه به نظر خبرگان، اگر سقف تعداد خوشه‌های نهایی را ۱۲ بدانیم، بر اساس قاعده فوق می‌توان کل داده‌های مورد خوشه‌بندی در فاز دوم (نورون‌های SOM) را برابر با $12^2 = 144$ تخمین زد. البته باید ذکر نمود که به کارگیری این روش برای یافتن تعداد نورون‌های اولیه SOM ابتکاری است و برخی پژوهش‌های دیگر نظیر پژوهش یو و لی (۲۰۱۳) تعداد خوشه اولیه را فقط بر مبنای نظر پژوهشگر محاسبه کرده است.



شکل ۴. خوشه‌های اولیه و تعداد اعضای آنها در گام اول SOM دو مرحله‌ای

شکل ۴، خروجی شبکه عصبی مصنوعی SOM را در قالب نورون‌های اولیه و تعداد اعضای هر یک به دست داده است. میزان فواصل مراکز این ۱۴۴ خوشه اولیه (نورون‌های شبکه) نسبت به یکدیگر یا همان شباهت‌های همسایگی در شکل ۵ به نمایش درآمده است. هرچه فاصله

بین دو نرون بیشتر باشد، خط همسایگی آنها تیره‌تر بوده و هرچه فاصله آنها کمتر باشد یعنی شباهت بیشتری به یکدیگر داشته باشند، با رنگ روشن‌تری به هم متصل شده‌اند. این شکل به ماتریس U^1 شهرت دارد. این ماتریس کمک می‌کند که داده‌های چندبعدی در قالب یک شکل تصویری دو بعدی قابل تفسیر شوند (اولتچ و سیمون، ۱۹۹۰). بر اساس نظر هامل و برون (۲۰۱۲)، تفسیر طرح ماتریس U و تشخیص خوشه‌ها در آن - به ویژه در ابعاد بالا - دشوار است. مزیت مهم روش این پژوهش آن است که شکل ۵ - ماتریس U - و مفاهیم موجود در آن با یک رویکرد ریاضی به تفسیر نهایی منجر می‌شود و صرفاً از دیدگاه بصری به منظور تحلیل آن استفاده نمی‌شود که وقوع خطا را محتمل سازد.



شکل ۵. شدت فواصل بین نرون‌های همسایه در شبکه SOM

فرآیند سنجش فاصله در آموزش شبکه عصبی SOM در این پژوهش بر اساس فاصله پیوندی (linkdist) با نرخ یادگیری پیش فرض نرم‌افزار و ۲۰۰ تکرار در آموزش به اتمام رسیده است.

با توجه به ۱۴۴ نرون فوق، فرآیند مرحله دوم خوشه‌بندی با روش K-means بر روی مراکز این نرون‌ها انجام شده است. در این مرحله مراکز خوشه‌های اولیه به ۵ تا ۱۲ تعداد خوشه جدید تقسیم شده و برای هر یک، شاخص‌های اعتبارسنجی خوشه‌بندی که در بخش پیش معرفی شدند محاسبه می‌گردد.

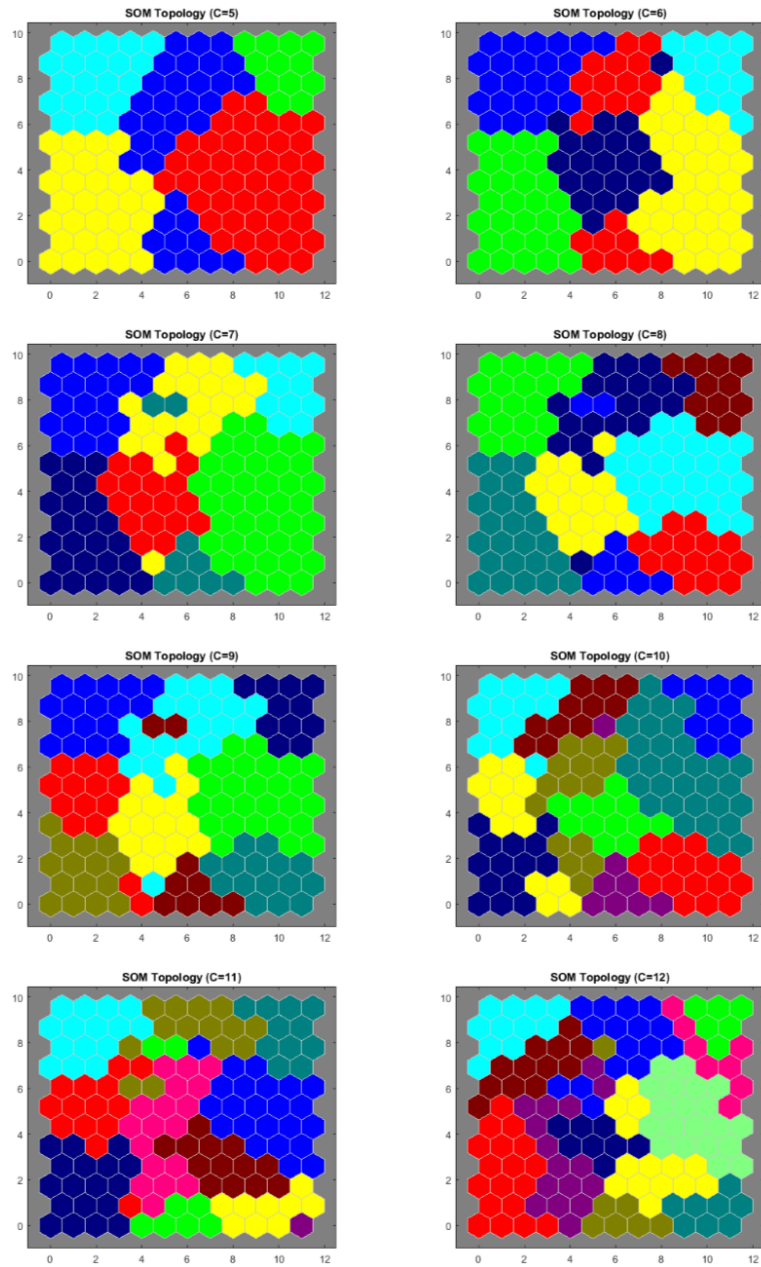
شکل ۶ خروجی نهایی حاصل از الگوریتم دو مرحله‌ای SOM را نمایش می‌دهد. هر یک از نمودارهای این شکل، در بردارنده خروجی روش K-means به منظور خوشه‌بندی مراکز خوشه‌های اولیه SOM است که برای تعداد مشخصی خوشه نهایی اجرا شده است. در این شکل، خوشه‌های هر ساختار با رنگ‌های مجزا نمایش داده شده است. به منظور ایجاد این شکل در نرم‌افزار متلب، در بخش تعیین رنگ تابع `plotsomtop` شرط‌های منطقی ایجاد شد تا هر خوشه ایجاد شده توسط K-means با رنگ مجزایی نسبت به دیگر خوشه‌ها تصویر گردد.

جدول ۱ نیز مقادیر مربوط به شاخص‌های اعتبارسنجی را برای هر یک از تعداد خوشه ۵ تا ۱۲ نشان می‌دهد که توسط نرم‌افزار متلب استخراج شده است.

جدول ۱. مقادیر شاخص‌های اعتبارسنجی برای هر یک از تعداد خوشه‌های نهایی

تعداد خوشه	Dunn	Silhouette	DB	VRC	RS
۵	۰/۰۶۲۱۲۶	۰/۴۹۱۴۳۲	۱/۰۴۴۰۴۷	۸۱/۱۴۵۱۸	۱۳/۳۷۴۳۹
۶	۰/۰۷۸۳۴۶	۰/۴۹۷۶۸۲	۰/۹۸۲۷	۸۱/۱۵۳۳۳	۱۱/۳۲۰۱۲
۷	۰/۰۷۳۶۶۲	۰/۴۹۸۱۵۶	۰/۹۴۵۴۵۷	۷۹/۰۱۰۳۲	۱۰/۰۰۰۴۶
۸	۰/۰۸۷۸۵۳	۰/۴۶۹۶۷۷	۱/۰۳۲۹۱۷	۷۷/۳۱۸۰۹	۸/۹۵۷۵۵۲
۹	۰/۰۸۷۸۵۳	۰/۴۲۲۵۸۱	۱/۰۸۸۸۳۱	۷۲/۷۶۷۶۲	۷/۸۶۳۵۷۹
۱۰	۰/۰۹۲۲۰۵	۰/۴۲۹۷۱۷	۱/۰۷۶۵۱۲	۶۹/۱۳۹۹۴	۶/۴۵۶۶۸۶
۱۱	۰/۱۲۹۱۷۹	۰/۴۵۲۳۱۶	۰/۹۷۱۲۰۳	۷۱/۵۴۱۲۵	۵/۴۱۲۱۷۹
۱۲	۰/۰۸۶۲۰۱	۰/۴۰۸۹۸۶	۱/۰۳۰۷۴۹	۶۵/۶۲۷۷۱	۴/۷۰۵۲۴۸

همان‌طور که از جدول ۱ روشن می‌شود، بهترین تعداد خوشه از منظر شاخص Dunn، Silhouette، DB، VRC و RS به ترتیب تعداد ۱۱ خوشه، ۷ خوشه، ۷ خوشه، ۶ خوشه و ۵ خوشه است. بر اساس متد رایج در برخی از پژوهش‌ها، در چنین شرایطی معمولاً ساختاری برگزیده می‌شود که بیشترین تعداد برتری را در شاخص‌ها داشته باشد که در اینجا تعداد ۷ با دو بار برتری حائز این شرایط است. رویکرد غیرجبرانی در این گونه روش‌ها باعث می‌شود که تأثیرات متقابل شاخص‌ها بر هم در نظر گرفته نشود، اما نوآوری این پژوهش مبنی بر بهره‌گیری از رویکرد جبرانی، این امکان را ایجاد می‌کند که نگاه دقیق‌تری نسبت به این موضوع فراهم شود. ممکن است وضعیت ساختاری در چندین شاخص مطلوب باشد، اما در شاخص‌های دیگر نامطلوب باشد و یا آنکه فاصله مطلوبیت آن در شاخص‌های برتری یافته، نسبت به دیگر ساختارها اندک باشد. برای پرهیز از تصمیم‌گیری مبهم در چنین شرایطی باید به روش‌های تصمیم‌گیری روی آورد تا بتوان بهترین نتیجه ممکن و قضاوت شدنی را برگزید. این موضوع همان رویکرد جبرانی را نسبت به گزینه‌ها و شاخص‌ها نشان می‌دهد. با استفاده از اطلاعات جدول ۱ و با بهره‌گیری از رویکرد تصمیم‌گیری چندشاخصه می‌توان قضاوت معتبری را پیرامون بهترین تعداد خوشه‌ها در بازه معین شده ارائه کرد. بدین منظور با در نظر گرفتن تعداد خوشه‌ها به عنوان گزینه و شاخص‌های اعتبارسنجی به عنوان شاخص‌های تصمیم‌گیری، فرآیند اجرای روش‌های MADM قابل انجام است.



شکل ۶. خروجی حاصل از SOM دو مرحله‌ای برای تعداد خوشه‌های نهایی ۵ تا ۱۲

در این پژوهش با استفاده از روش TOPSIS نسبت به اولویت‌بندی و بهینه‌یابی گزینه‌ها اقدام شده است. در روش TOPSIS ساختار ایده‌آل و ساختار ضدایده‌آل به ترتیب بر اساس بهترین و بدترین مقادیر در هر شاخص فراهم شده (با مد نظر قرار دادن مثبت یا منفی بودن شاخص مورد نظر) و فاصله هر یک از ساختارها با این دو ساختار سنجیده می‌شود تا در نهایت بهترین ساختار که بیشترین فاصله از ضدایده‌آل و کمترین فاصله از ایده‌آل را دارد، تشخیص داده شود (هوآنگ و یون، ۱۹۸۱). این الگوریتم نیاز به وزن‌دهی به شاخص‌ها دارد که روش‌های متعددی برای آن توسعه یافته است. در این پژوهش به منظور سهولت DM فرآیند وزن‌دهی شاخص‌ها با در نظر گرفتن وزن‌های مساوی برقرار شده است. با توجه به حذف نظرات خبرگان از فرآیند خوشه‌بندی، در فرآیند وزن‌دهی به شاخص‌ها نیز نظرات خبرگان حذف شده تا استقلال پژوهش از دخالت DM حفظ شود، از این روی، استفاده از وزن‌های یکسان باعث می‌شود که حتی از منظر وزن‌دهی و اهمیت‌سنجی شاخص‌های اعتبارسنجی نیز دخالت انسانی وجود نداشته باشد و همچنین از تأثیرپذیری از میزان پراکندگی آنها جلوگیری شود. بر این اساس، با استفاده از الگوریتم TOPSIS، فرآیند انتخاب بهترین ساختار - تعداد مطلوب برای خوشه‌های نهایی - اجرا شد. نتیجه اجرای این الگوریتم در نرم‌افزار متلب در جدول ۲ درج شده است.

جدول ۲. امتیاز هر یک از ساختارها در روش TOPSIS

امتیاز TOPSIS	تعداد خوشه
۰/۵۷۰۱	۵
۰/۵۷۰۴	۶
۰/۴۸۰۷	۷
۰/۴۶۳۹	۸
۰/۳۶۶۱	۹
۰/۳۰۲۵	۱۰
۰/۴۶۱۸	۱۱
۰/۱۹۷۷	۱۲

بر اساس جدول ۲، بهترین ساختار با تعداد ۶ خوشه بوده که بالاترین امتیاز را از رویکرد تصمیم گیری چندشاخصه اخذ کرده است. با این حساب در شکل ۶ نیز ساختار $C=6$ نشان دهنده بهترین ساختار ممکن در بازه تعداد خوشه‌های معین شده است. از این روی، کل فروشگاه‌های آنلاین به ۶ خوشه نهایی تقسیم شده‌اند.

مقادیر مربوط به مراکز این شش خوشه در جدول ۳ به نمایش درآمده است. مقادیر مذکور در قالب اعداد نرمالایز شده - که خوشه‌بندی بر اساس آنها انجام شد - نشان داده شده است. دلیل این موضوع آن است که گزارش مقادیر دقیق مربوط به فروش و زمان با سیاست‌های شرکت تأمین کننده سازگار نبوده و لذا به دلیل عدم مجوز شرکت تأمین کننده در این مقاله به صورت شفاف ارائه نشده است. با این وجود و بر اساس مقایسه همین مقادیر، تحلیل هر یک از خوشه‌ها در ادامه می‌آید.

جدول ۳. مراکز شش خوشه نهایی حاصل از مرحله دوم الگوریتم

خوشه	عمر	آخرین فعالیت	آخرین فروش	مدت فعالیت	فروش ماه اخیر	رشد فروش ماه اخیر	فروش هفته اخیر	رشد فروش هفته اخیر
یکم	۰/۰۵	۰/۳۷	۰/۱۰	۰/۱۳	۰	۰	$۲/۸۶ \times ۱۰^{-۳}$	$-۲/۷۵ \times ۱۰^{-۲}$
دوم	۰/۱۳	۰/۴۴	۰/۶۱	۰/۱۱	$\times ۱۰^{-۶}$ ۶/۸۷	$۵/۵۹ \times ۱۰^{-۶}$	$۲/۷۸ \times ۱۰^{-۷}$	$۱/۷۵ \times ۱۰^{-۵}$
سوم	۰/۶۰	۰/۷۸	۰/۰۳	۰/۰۴	۰	۰	$۱/۷۲ \times ۱۰^{-۶}$	$۹/۱۰ \times ۱۰^{-۵}$
چهارم	۰/۱۱	۰/۸۸	۰/۸۶	۰/۵۲	$\times ۱۰^{-۳}$ ۸/۱۵	$۷/۶۵ \times ۱۰^{-۳}$	$۹/۴۳ \times ۱۰^{-۵}$	$-۲/۶۰ \times ۱۰^{-۴}$
پنجم	۰/۵۹	۰/۸۵	۰/۸۱	۰/۱۲	$\times ۱۰^{-۵}$ ۴/۷۵	$۶/۵۵ \times ۱۰^{-۵}$	$۱/۹۳ \times ۱۰^{-۵}$	$-۶/۳۰ \times ۱۰^{-۴}$
ششم	۰/۱۰	۰/۷۱	۰/۲۹	۰/۳۵	۰	۰	$۱/۱۱ \times ۱۰^{-۶}$	$۱/۷۳ \times ۱۰^{-۵}$

خوشه یک به عنوان فروشگاه‌های تازه تأسیس تلقی می‌شود که عمر بسیار کمتری نسبت به خوشه‌های دیگر دارد اما در همین عمر کم توانسته است فروش قابل توجهی کسب کند. بدیهی است که سیاست حمایتی از این خوشه می‌تواند در پیش‌برد استراتژی‌های گسترش بازار تأمین‌کننده مؤثر باشد. وضعیت خوشه‌های دو، چهار و شش تا حدودی به یکدیگر شبیه است چرا که هر سه در زمره فروشگاه‌های میان‌سال طبقه‌بندی می‌شوند. اما میزان فعالیت در خوشه‌های دوم و چهارم به مراتب از خوشه ششم بیشتر است که این موضوع به حاشیه‌رانده شدن خوشه ششم را تقویت می‌کند. با این حال، فروش ماهانه و هفتگی فروشگاه‌های خوشه چهارم به مراتب از خوشه دوم بالاتر است و این نشان‌گر پیشتازی در فروش در عین عمری نسبتاً میانه است. همچنین میزان فعالیت‌های اخیر این خوشه بیان‌گر به روز بودن و جدی‌تر بودن فروشگاه‌های موجود در آن برای رقابت در این بازار است. به نظر می‌رسد که نبض فروش در دست این فروشگاه‌هاست و بیشترین تعامل تأمین‌کننده در بستر وب با این کاربران انجام می‌شود که این خود اهمیت تداوم همکاری و سیاست‌های تشویقی و تخفیفی تأمین‌کننده را نسبت به این خوشه روشن می‌کند. خوشه‌های سه و پنج بیشترین سن تقویمی را دارند و این در حالی است که فروش فروشگاه‌های خوشه سه در ماه اخیر قابل توجه نیست و این موضوع نشان‌دهنده رکود جدی در این خوشه است. از سوی دیگر فروشگاه‌های خوشه پنجم اگرچه بیشتری فروش را ندارند اما به نظر می‌رسد در یک مسیر پایدار فروش قرار گرفته‌اند چراکه همچنان فعالیتی مشابه خوشه چهارم (شاخص‌های آخرین فعالیت و آخرین فروش) دارند و این یعنی جریان فروش همچنان در آنها برقرار است و نیاز به سیاست‌های حمایتی و پشتوانه‌ای دارند. در مقابل، خوشه سوم قابل حذف است و می‌توان با رویکرد اختصارگونه نسبت به حذف منابع مجازی که در اختیار آنان است، اقدام کرد تا از هدررفت منابع جلوگیری به عمل آید.

نتیجه‌گیری و پیشنهادها

در این پژوهش، فروشگاه‌های آنلاین با استفاده از شاخص‌های تعریف شده توسط تأمین‌کننده آنها با الگوریتم دو مرحله‌ای SOM دسته‌بندی شده و مورد تحلیل خوشه‌ای واقع شدند. در بسیاری از پژوهش‌ها کمترین تأکید بر تعداد خوشه‌هاست در حالی که این مسأله در مواردی چون تخصیص منابع، برنامه‌ریزی و تنوع سیاست‌های اتخاذ شده تأثیر خواهد داشت.

بهره‌گیری از خوشه‌بندی با این الگوریتم این امکان را ایجاد می‌کند که تفسیر ماتریس U حاصل از شبکه نگاشت خودسازمان‌ده با سهولت بیشتری صورت گیرد. از سوی دیگر، در این پژوهش تعداد خوشه‌ها فقط بر اساس یک شاخص اعتبارسنجی خوشه‌بندی انجام نشد، بلکه به منظور ایجاد اتقان و استواری بیشتر در این زمینه، از پنج شاخص به طور همزمان بهره گرفته شد. با توجه به نتایج متفاوت هر یک از شاخص‌ها، با بهره‌گیری از رویکرد تصمیم‌گیری چندشاخصه و وزن‌دهی شاخص‌ها این امکان فراهم آمد که بهترین ساختار خوشه‌بندی بر اساس بیشترین امتیاز از مجموع تمامی شاخص‌ها انتخاب گردد. چنین امکانی موجب خواهد شد که قضاوت مستدل‌تری پیرامون انتخاب تعداد خوشه‌ها صورت گیرد و مانند بسیاری از پژوهش‌های مرسوم صرفاً با تخمین و برآورد خبرگان یا اتکا به یک یا چند شاخص با رویکرد غیرجبرانی، این کار صورت نپذیرد. مانند پژوهشی که یو و لی (۲۰۱۳) انجام داده و پس از خوشه‌بندی با SOM فقط از یک شاخص برای یافتن تعداد بهینه خوشه‌ها استفاده کرده‌اند. در پژوهش‌های دیگری نیز چون مقاله حسین مرشدی و معماریان (۱۳۹۱) نیز اگرچه از چندشاخص استفاده شده اما به دلیل رویکرد غیرجبرانی تأثیرات متقابل را در نظر گرفته نشده است.

به منظور توسعه این پژوهش، بهره‌گیری از یک رویکرد تحلیلی توصیه می‌شود. استفاده از رویکرد SOM با ساختارهای اولیه متفاوت و مقایسه آن با پاسخ نهایی موجب تدقیق افزون‌تر جواب خواهد شد. این رویکرد کمک می‌کند که حساسیت پاسخ نهایی الگوریتم دو مرحله‌ای SOM را نسبت به ساختار اولیه شبکه دریابیم و در نتیجه بهترین ساختار شدنی را پیدا کنیم. علاوه بر این موضوع، با توجه به محدودیت‌هایی که در پایگاه داده تأمین‌کننده وجود داشته است، در صورت رفع این محدودیت‌ها و استفاده از شاخص‌های گسترده‌تر نیز می‌تواند به غنای این پژوهش یاری رساند.

منابع

- Calinski, T., & Harabasz, J. (1974). *A dendrite method for cluster analysis*. Communications in Statistics, ۳(۱): ۱-۲۷.
- Chiang, M. M. & Mirkin, B. (2010). Intelligent choice of the number of clusters in kmeans clustering: an experimental study with different cluster spreads. Journal of Classification, ۲۷(۱): ۳-۴۰.
- Clark, S., Sarlin, P., Sharma, A. & Sisson, S. A. (2015). *Increasing dependence on foreign water resources? An assessment of trends in global virtual water flows using a self-organizing time map*. Ecological Informatics, ۲۶(۲): ۱۹۲-۲۰۲.
- Davies, D. L., & Bouldin, D. W. (1979). *A Cluster Separation Measure*. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-۱(۲): ۲۲۴-۲۲۷.
- Dunn, J. C. (1974). Well separated clusters and optimal fuzzy partitions. Journal of Cybernetica, ۴(۱): ۹۵-۱۰۴.
- Ghazanfari, M., Alizadeh, S. & Teimourpour, B. (2008). *Data Mining & Knowledge Discovery*. Tehran: Iran University of Science & Technology (IUST) Press. (in persian)
- Hamel, L. & Brown, C. W. (2012). *Improved Interpretability of the Unified Distance Matrix with Connected Components*. Proceeding of the 7th International Conference on Data Mining (DMIN'11): Las Vegas Nevada, USA.
- Hosseini Morshedy, A. & Memarian, H. (2012). *Zoning of RQD Parameter, Based on Faults and Self-Organizing Map in Semilan Dam Site*. GEOSCIENCES, ۲۱(۸۴), ۹۹-۱۱۲.
- Hu, W. & Jing, Z. (2008). *Study of Customer Segmentation for Auto Services Companies Based on RFM Model*. The International Conference on Innovation Management, December ۱۰-۱۱.
- Hung, C. & Tsai, C. F. (2008). *Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand*. Expert Systems with Applications, ۳۴(۱): ۷۸۰-۷۸۷.
- Hwang, C.L., & Yoon, K. (1981). *Multiple Attribute decision making*. Springer.

Kalyani, S. & Swarup, K. S. (2011). *Particle swarm optimization based K-means clustering approach for security assessment in power systems*. Expert Systems with Applications, ۳۸(۹), ۱۰۸۳۹-۱۰۸۴۶.

Khadivar, A., Razmi, Z. & Hamed, P. (2014). *Customer clustering for appointing rebating strategies, case study: Kadbano Co*. New Marketing Research, ۳(۳), ۱۳۵-۱۵۰.

Kim, K. J. & Ahn, H. (2008). *A recommender system using GA K-means clustering in an online shopping market*. Expert Systems with Applications, ۳۴(۲): ۱۲۰۰-۱۲۰۹.

Kohonen, T. (1989). *Self-Organization and Associative Memory*. Berlin: Springer.

Kohonen, T. (1998). *The self-organizing map*. Neurocomputing, 21(1): 1-6.

Kotler, P. H. & Armstrong, G. (2013). *Principles of Marketing*, 14th Edition, New Jersey: Prentice Hall.

Liang, Y. H. (2010). *Integration of data mining technologies to analyze customer value for the automotive maintenance industry*. Expert Systems with Applications, 37(12): 7489-7496.

Lin, Y., Lee, P., Chang, T. & Ting, H. (2008). *Multi attribute group decision making model under the condition of uncertain information*. Automation in construction, ۱۷(۶), ۷۹۲-۷۹۷.

López, J. J., Aguado, J. A., Martín, F., Muñoz, F., Rodríguez, A. & Ruiz, J. E. (2011). *Hopfield-K-Means clustering algorithm: A proposal for the segmentation of electricity customers*. Electric Power Systems Research, 81(2):716-724.

Momeni, M. (2011). *Data Clustering: Cluster Analysis*. Tehran: Mansoor Momeni. (in persian)

Redmond, S., & Heneghan, C. (2007). *A method for initialising the K-means clustering algorithm using KD-trees*. Pattern Recognition Letters, ۲۸(۸), ۹۶۵-۹۷۳.

Rouseeuw, P. J. (1987). *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis*. Journal of Computational and Applied Mathematics, ۲۰(۱): ۵۳-۶۵.