



بهبود الگوریتم خوشه بندی مشتریان برای توزیع قطعات یدکی با رویکرد داده کاوی (k-means)

محمد مهدی سپهری* و مهرداد کارگری

چکیده:

کلمات کلیدی

خوشه بندی مشتریان با رویکرد داده کاوی برای بهینه سازی برنامه حمل و نقل یکی از مباحث مطرح در حوزه مدیریت ارتباط با مشتریان است. در این مقاله داده‌ها و اطلاعات واقعی توزیع قطعات یدکی شرکت ایساکو در طی ۳۶ ماه گذشته مورد بررسی قرار گرفته است و به کمک ابزار داده کاوی شباهت رفتار تقاضای قطعات یدکی توسط مشتریان در مناطق مختلف کشور ایران سنجیده شده است. تابع سنجش شباهت رفتار مشتریان در مناطق مختلف براساس ترکیب قواعد تعریف گردیده است. براساس این تابع و به کمک الگوریتم k-means عملیات خوشه بندی انجام گرفته و مشتریان مشابه در یک خوشه قرار گرفته‌اند. تابع سنجش شباهت رفتار مشتریان در پنج مرحله تشکیل شده است. تابع شباهت مشتریان براساس فواصل اقلیدسی (مکان استقرار آنها)، زمان تقاضای مشتریان و مقدار ارزش حجمی تقاضای مشتریان به صورت جداگانه تعریف شده است. سپس این سه عامل با هم ترکیب شده و تابع DCB به وجود آمده است. در مرحله نهایی با در نظر گرفتن وزن‌های مختلف برای سال‌ها و فصول گوناگون تابع BCD بهبود داده شده است. عملیات خوشه بندی به وسیله سه تابع فاصله اقلیدسی، تابع DCB و تابع BCD بهبود یافته در نرم افزار R انجام و تابع BCD بهبود یافته به عنوان بهترین تابع برای خوشه بندی انتخاب شده است. سپس با استفاده از مدل رفتار هر بخش تحلیل شده و سیاست‌های توزیع مناسب برای آن بخش تبیین شده است. نتایج حاصل بیانگر کاهش ۳۲ درصد هزینه‌های توزیع در شرکت ایساکو می‌باشد.

الگوریتم خوشه بندی مشتریان،
مدیریت ارتباط با مشتریان،
داده کاوی (k-means)،
تابع شباهت،
توزیع قطعات یدکی

۱. مقدمه

خوشه بندی مشتریان براساس مناطق جغرافیایی یکی از مباحث مطرح در حوزه مدیریت ارتباط با مشتریان است. خوشه بندی در واقع شکستن جمعیت زیاد مشتریان به بخش‌های مختلف است. بطوریکه مشتریان موجود در هر بخش به یکدیگر شبیه و مشتریان بخش‌های مختلف با یکدیگر متفاوت هستند. خوشه بندی

دیدگاهی کلی و سطح بالا از تمام بانک اطلاعاتی مشتریان ارائه داده و به صاحبان شرکت‌های توزیع امکان اعمال رفتار و سیاست‌های متفاوت و مناسب به مشتریان هر بخش را می‌دهد. در حالت ایده آل هر سازمان باید هر یک از مشتریان را بطور کامل بشناسد، ولی این کار امکان پذیر نیست و در واقع خوشه بندی این امکان را فراهم می‌آورد تا مشتریان که شبیه به هم هستند در یک بخش قرار گیرند. معیار شباهت براساس نوع صنعت و کسب و کار متفاوت است. در اینصورت مدیریت و شناخت این بخش‌ها بسیار ساده تر از شناخت تک تک مشتریان است. خوشه بندی به شیوه‌های مختلفی صورت می‌گیرد. در برخی تحقیقات صورت گرفته در این حوزه، با در نظر گرفتن سهم سود ایجاد شده، سود بالقوه و تعریف سود آوری مشتری یک مدل LTV پیشنهاد شده

تاریخ وصول: ۸۹/۶/۹

تاریخ تصویب: ۹۰/۴/۷

*تویسنده مسئول مقاله: دکتر محمد مهدی سپهری، عضو هیئت علمی
دانشگاه تربیت مدرس، mehdi.sepehri@modares.ac.ir
مهرداد کارگری، دانشجوی دکتری دانشگاه تربیت مدرس،
M_kargari@modares.ac.ir

در قسمت دوم روش خوشه بندی مشتریان بیان شده است. این قسمت شامل آماده سازی داده‌ها، معرفی توابع شباهت، الگوریتم خوشه بندی k -means براساس توابع شباهت، ارزیابی کیفیت خوشه بندی و بهبود توابع خوشه بندی بصورت ترکیبی می‌باشد. قسمت سوم مدل DTM برای ارزیابی خوشه ها را بیان میکند. مطالعه موردی در قسمت چهارم بیان شده است. این قسمت شامل مقایسه میزان تراکم در حالات مختلف، کیفیت خوشه ها در حالات مختلف و تحلیل خوشه ها در مدل DTM می‌باشد. نتایج و تحقیقات آینده در قسمت پنجم بیان شده است.

۲. متدولوژی بخش بندی مشتریان

در این مقاله ابتدا مشتریان بخش بندی می‌شوند و پس از ارزیابی و بهبود تابع فاصله، سیاست‌های رفتاری مناسب برای هر بخش تبیین می‌گردد. اولین مرحله مانند هر پروژه داده کاوی دیگر آماده سازی داده‌هاست. سپس تابع DCB برای سنجش تفاوت رفتار مشتریان براساس مفاهیم قواعد انجمن تعریف شده است و با جایگزینی این تابع در k -means خوشه بندی صورت گرفته است. بهبود خوشه بندی صورت گرفته نسبت به عملکرد تابع فاصله اقلیدسی در الگوریتم k -means با استفاده از تابع سنجش تراکم خوشه‌ها نشان داده شده است. تعداد خوشه‌ها با استفاده از تابع سنجش کیفیت خوشه بندی تعیین گردیده است. مراحل ذکر شده در ادامه به طور کامل تشریح شده‌اند.

۲-۱. آماده سازی داده‌ها

در این مرحله داده‌های بانک اطلاعاتی با توجه به اهداف و الگوریتم‌های بکار گرفته شده آماده سازی شده‌اند. مفاهیم پارامترهای مورد استفاده در تابع شباهت عبارتند از:

G : مجموعه همه گروه قطعات یدکی شرکت (حجم و ابعاد یکسان)
 T^C : بانک اطلاعات مشتریان

f : یک رکورد از جدول T^C است، هر رکورد این جدول شامل کد مشتری، نام مشتری، مختصات جغرافیایی مشتری، نوع و گروه کالای درخواستی، میزان تقاضا، شهر مشتری، زمان تقاضا، ارزش تقاضا و حجم تقاضا می‌باشد.

C_i : شهر i

g_{ia} : گروه کالای مورد نیاز a در شهر i

m_{ia} : مجموع ارزش حجمی کالای درخواستی گروه a در شهر i

برای مشاهده رفتار مشتریان نیاز به ارزیابی اطلاعات همه مشتریان و کل ارزش حجمی گروه کالای درخواستی توسط آن‌ها در طول پانزده سال گذشته است.

$Citycode_i$: کد شهر i

$goodgorup_i$: مجموعه گروه کالاهایی مورد نیاز شهر i

است و براساس ارزش فعلی، ارزش بالقوه و میزان وفاداری، مشتریان بخش بندی شده‌اند. [۵] سایون و همکاران در سال ۲۰۰۶ با در نظر گرفتن اهمیت شناخت مشتری در جهت ایجاد ارتباط بلندمدت، کسب وفاداری و سودآوری بیشتر مشتری، ارزش هر مشتری را تعیین نموده و سپس با در نظر گرفتن ارزش مشتریان، آن‌ها را بخش بندی نموده و استراتژی‌های مناسب برای هر بخش را تبیین نموده‌اند [۶].

تسای و چپو در سال ۲۰۰۴ به بسط متدولوژی جدید برای بخش بندی بازار بر مبنای متغیرهای خاص از جمله اقلام خریداری شده و میزان درآمد مرتبط با آن‌ها با در نظر گرفتن تراکنش‌های قبلی مشتریان پرداخته و بدین ترتیب بازار را بخش بندی نمودند [۷].

در برخی مطالعات صورت گرفته در حوزه بخش بندی بازار از روش‌های خوشه‌بندی مدل k -means, FUZZY k -means, SOM برای بخش بندی مشتریان استفاده شده است [۸].

همچنین در دیگر تحقیقات مشتریان، براساس معیار تازگی، تکرار و ارزش پولی به گروه‌های همگن بخش بندی شده‌اند، سپس سیاست‌های بهینه بازار بای برای هر بخش تدوین شده است [۹].

همانطور که اشاره شد بخش بندی مشتریان به شیوه‌های گوناگون صورت می‌گیرد. در شرکت ایساکو میزان شباهت رفتار مشتریان با توجه به نوع داده‌ها و اطلاعات موجود از مشتریان به محل استقرار آن‌ها، شهر مشتری، گروه کالای مورد نیاز آن‌ها، میزان تقاضای آن‌ها و زمان تقاضای آن‌ها وابسته است. بر این اساس در این مقاله از معیارهای شهر، فاصله اقلیدسی، گروه کالای مورد نیاز، میزان و زمان تقاضا برای استخراج تابع شباهت استفاده شده است. ابتدا دو معیار گروه کالای مورد نیاز و حجم تقاضا با استفاده از الگوریتم k -means برای خوشه بندی مشتریان استفاده شده است. سپس تابع فاصله اقلیدسی به صورت جداگانه در الگوریتم k -means برای خوشه بندی مشتریان منظور شده است. برای بهبود تابع شباهت معیارهای شهر، فاصله اقلیدسی، گروه کالای مورد نیاز، میزان و زمان تقاضا با هم ترکیب شده و مجدداً با الگوریتم k -means برای خوشه بندی مشتریان بکار رفته است. پس از تعیین بهترین خوشه بندی براساس معیار تراکم خوشه بندی، بهترین تعداد خوشه با در نظر گرفتن معیار سنجش کیفیت خوشه بندی محاسبه شده است.

هدف این مقاله بهینه سازی برنامه توزیع قطعات یدکی شرکت ایران خودرو با رویکرد داده کاوی و خوشه بندی فروشگاه های مجاز در نقاط مختلف کشور میباشد. برای این منظور از سه تابع شباهت فاصله اقلیدسی، میزان تقاضای مشتریان و زمان تقاضای مشتریان بصورت همزمان استفاده شده است. سپس با در نظر گرفتن وزن‌های مختلف برای فصول سال تابع شباهت ترکیبی بهبود داده شده است. ساختار این مقاله در ادامه بشرح زیر است.

همچنین اگر رکورد t_j^c برای مشتریان شهر j شامل $(Citycode_j, goodgorup_j, moneyset_j)$ باشد و گروه کالا و ارزش حجمی هر گروه کالا به ترتیب برابر $g_{j1}, g_{j2}, \dots, g_{jn}$ و $m_{j1}, m_{j2}, \dots, m_{jn}$ باشد شباهت دو شهر i و j بر مبنای تقاضای مشتریان دو شهر به کمک رابطه (۳) زیر محاسبه می شود.

$$\left[\begin{array}{l} sim(C_i, C_j) = \frac{\sum_{a=1}^s \sum_{b=1}^s m_{ia} \times m_{jb} \times \delta(g_{ia}, g_{jb})}{\sum_{a=1}^s \sum_{b=1}^s m_{ia} \times m_{jb}} \quad C_i \neq C_j \\ sim(C_i, C_j) = 1 \quad C_i = C_j \end{array} \right] \quad (3)$$

متعاقباً فاصله دو شهر C_i ، C_j یا تفاوت رفتار تقاضای مشتریان دو شهر به کمک رابطه (۴) محاسبه می شود.

$$\left[\begin{array}{l} DCB(C_i, C_j) = 1 - sim(C_i, C_j) \quad C_i \neq C_j \\ DCB(C_i, C_j) = 0 \quad C_i = C_j \end{array} \right] \quad (4)$$

دو معادله بالا علاوه بر در نظر گرفتن گروه کالای مورد نیاز مشتریان و ارزش حجمی تقاضای مشتریان هر شهر در میزان کاهش هزینه های حمل و نقل شرکت ایساکو تاثیرگذار است. برای مثال اگر گروه کالای A معمولاً با گروه کالای B ارسال شود و کمتر با گروه کالای C ارسال شود، وابستگی تقاضای دو گروه کالای A, B باید بیشتر از دو گروه کالای A, C باشد. همچنین نادیده گرفتن این وابستگی ها و یکسان گرفتن رفتار همه گروه کالاها، باعث ایجاد شباهت یک طرفه می شود. در صورتی که باید برای کامل شدن محاسبات از شباهت دو طرفه استفاده نمود. به این اساس در این مقاله از شباهت دو طرفه گروه کالاها استفاده شده است.

۳-۲. خوشه بندی

خوشه بندی وظیفه تقسیم یک گروه ناهمگن را به چندین زیر گروه که به اصطلاح به آن خوشه می گویند را به عهده دارد. در این تحقیق از الگوریتم k -means برای خوشه بندی مشتریان استفاده شده است. برای سنجش فاصله مشتریان دو شهر در الگوریتم k -means از تابع DCB ، تابع فاصله اقلیدسی و سپس با ترکیب این دو تابع و تاثیر داده زمان سفارش مشتریان تابع DCB بهبود داده شده است.

$$goodgorup_i = \{ g_{ia} | g_{ia} \in G \}$$

$moneyset_i$: مجموعه ارزش حجمی کالای درخواستی مشتریان شهر i

$$moneyset_i = \{ m_{ia} | a=1, 2, \dots, |goodgorup_i| \}$$

$GISplan_i$: مختصات جغرافیایی محل استقرار مشتری

$Demandtime_i$: زمان و تاریخ سفارش مشتری

بنابراین کالای درخواستی مشتریان شهر C_i را می توان به صورت t_i^c نشان داده و در بانک اطلاعات مشتریان (T^c) ذخیره نمود.

$$T^c_i = \{ Citycode_i, goodgorup_i, moneyset_i, Location plan_i, Demandtime_i \}$$

۲-۲. تابع فاصله تفاوت رفتار تقاضای مشتریان (DCB)

برای اندازه گیری تفاوت رفتار مشتریان در هر شهر، ابتدا وابستگی هر گروه کالا با گروه کالاهای دیگر بر اساس مفهوم پشتیبانی در قواعد انجمن محاسبه شده است. برای محاسبه این تفاوت از رابطه (۱) استفاده شده است.

$$\| t^c \in T^c_i / t^c \text{ Contains } \{ g_{ia}, g_{ib} \} \| \quad (1)$$

$$S(g_{ia}, g_{ib}) = \frac{\| t^c \|}{\| T^c_i \|}$$

$S(g_{ia}, g_{ib})$: نسبت تعداد تراکنش هایی که در آن دو گروه کالای a, b از شهر i وجود دارد به کل تعداد تراکنش های شهر i را نشان می دهد. اگر دو گروه کالا به ندرت با هم سفارش دهند در نتیجه مقدار پشتیبانی آن ها خیلی کم خواهد شد [۲]. در تابع DCB فاصله بر اساس میزان شباهت گروه کالا می باشد و نیاز به در نظر گرفتن وابستگی دوطرفه انواع گروه قطعات یدکی دارد. لذا برای محاسبه وابستگی دو طرفه گروه قطعات یدکی ضریب (۷) به کمک رابطه (۲) محاسبه می نمایم [۱۳].

$$S \{ g_{ia}, g_{ib} \} \quad (2)$$

$$\gamma(g_{ia}, g_{ib}) = \frac{S(g_{ia}, g_{ib})}{S(g_{ia}) + S(g_{ib}) - S(g_{ia}, g_{ib})}$$

مقدار $\gamma(g_{ia}, g_{ib})$ بین صفر و یک است. اگر $g_{ia} = g_{ib}$ باشد آنگاه $\gamma(g_{ia}, g_{ib}) = 1$ می باشد. حال میزان شباهت تقاضای مشتریان دو شهر به صورت زیر ارزیابی می شود.

اگر رکورد t_i^c برای مشتریان شهر i شامل $(Citycode_i, goodgorup_i, moneyset_i)$ باشد و گروه کالا و ارزش حجمی هر گروه کالا به ترتیب برابر $g_{i1}, g_{i2}, \dots, g_{in}$ و $m_{i1}, m_{i2}, \dots, m_{in}$ باشد

محل استقرار مشتریان شهر زیاد باشد. فاصله یا عدم شباهت مشتریان دو شهر با استفاده فاصله اقلیدسی به کمک رابطه (۵) محاسبه می‌شود.

$$Dis(C_i, C_j) = \sqrt{\sum_{l=1}^L [(X_{il} - X_{jl})^2 + (Y_{il} - Y_{jl})^2]} \quad (5)$$

در این تحقیق ابتدا تابع فاصله در الگوریتم k-means مطابق تابع فاصله اقلیدسی به صورت معادله فوق در نظر گرفته شده است و سپس تابع DCB به دو صورت ساده و ترکیبی (بهبود یافته) در نظر گرفته شده است.

۲-۵. الگوریتم خوشه بندی k-means براساس تابع DCB

در این خوشه بندی در الگوریتم k-means برای محاسبه فاصله از معادله (۴) استفاده شده است. فاصله مشتریان هر شهر به صورت دویبدو محاسبه شده است و در ماتریس به شکل زیر قرار گرفته است. در این ماتریس n تعداد مشتریان است، d_{ij} نشان دهنده فاصله بین مشتریان C_i و C_j و $(i, j) \leq n$ است. باتوجه به معادله (۴) قطر اصلی این ماتریس یعنی فاصله هر مشتری با خودش صفر است.

$$Dis = \begin{bmatrix} 0 & d_{i2} & d_{ij} & d_{1n} \\ d_{21} & 0 & d_{2j} & d_{2n} \\ d_{i1} & d_{i2} & 0 & d_{in} \\ d_{n1} & d_{n2} & d_{nj} & 0 \end{bmatrix} \quad (6)$$

این ماتریس به عنوان ورودی به الگوریتم k-means در نرم افزار R داده شده است. در واقع تابع فواصل یا عدم شباهت بین مشتریان یکبار براساس تابع فاصله اقلیدسی، یکبار براساس تابع فاصله DCB و یکبار هم بر اساس DCB بهبود یافته محاسبه شده است.

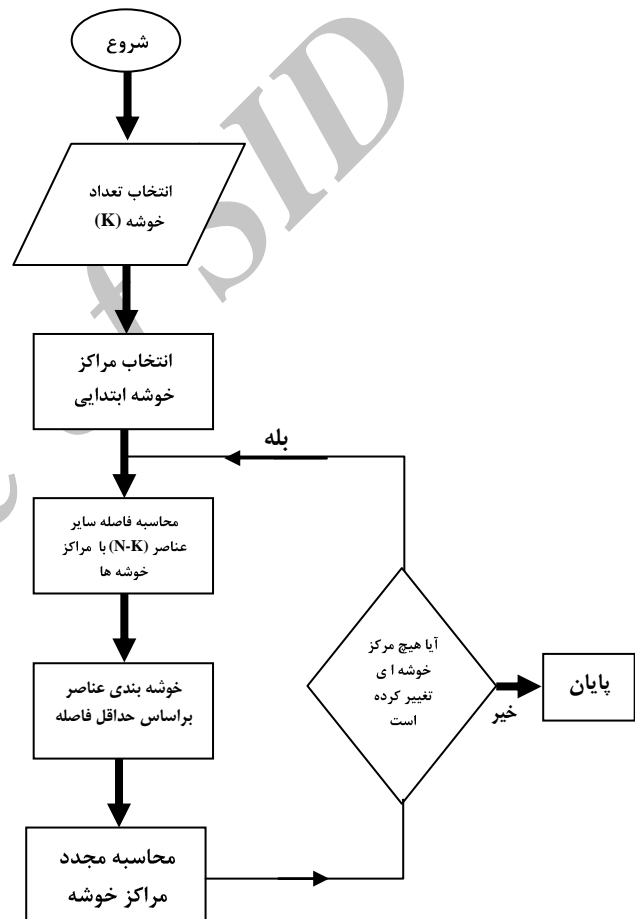
۲-۶. ارزیابی کیفیت خوشه بندی

معیارهای زیادی برای محاسبه کیفیت خوشه‌های حاصل از خوشه بندی وجود دارد. در خوشه بندی مطلوب باید شباهت درون خوشه‌ها حداکثر و عدم شباهت بین خوشه‌ها حداکثر شود. کلیه معیارهای سنجش کیفیت مبتنی بر این دو مفهوم هستند. برخی از آن‌ها فقط می‌توانند همگنی درون خوشه‌ای یا نا همگنی برون خوشه‌ای را تضمین کنند ولی برخی دیگر هر دو عامل را به اهم در نظر می‌گیرند [۱۳].

۲-۴. الگوریتم خوشه بندی k-means براساس تابع فاصله

اقلیدسی

در الگوریتم k-means ابتدا K عضو به صورت تصادفی از میان N عضو به عنوان مراکز خوشه‌ها انتخاب می‌شود و سپس N-K عضو باقی مانده به نزدیک‌ترین خوشه تخصیص می‌یابند. بعد از تخصیص همه اعضا مراکز خوشه مجدداً محاسبه می‌شوند و این کار با توجه به مراکز جدید به خوشه‌ها تخصیص می‌یابند و این کار تا زمانی که مراکز هر خوشه ثابت بمانند ادامه می‌یابد خلاصه این مراحل در نمودار (۱) نشان داده شده است [۲].



شکل ۱. الگوریتم خوشه بندی k-means

اگر رکورد i برای مشتریان شهر C_i شامل و

$(Citycode_i, Location plan_i, Demondtime_i)$ و

$Location plan =$

$[(X_{i1}, Y_{i1}), (X_{i2}, Y_{i2}), \dots, (X_{iL}, Y_{iL})]$

محل استقرار مشتریان شهر i باشد و رکورد j برای مشتریان

شهر C_j شامل

$(Citycode_j, Location plan_j, Demondtime_j)$

و

$Location plan_i, \dots, (X_{jL}, Y_{jL}) = [(X_{j1}, Y_{j1}), (X_{j2}, Y_{j2})]$

$$\delta_{nm} = \text{sim}(C^n, C^m) \quad (۱۲)$$

معادله (۱۰)، μ_n را به عنوان میانگین شباهت بین مرکز خوشه C^n و همه مشتریان خوشه O^n تعریف می کند.

معادله (۱۱) توضیح می دهد که μ_m میانگین شباهت بین مرکز خوشه C^m و همه مشتریان در خوشه O^m است. معادله (۱۲) δ_{nm} را به عنوان شباهت C^n و C^m تعریف می کند.

مقدار $p(k)$ با توجه به تابع کیفیت خوشه بندی در معادله (8)، برای یک محدوده پیشنهادی (حد بالای t و حد پائین s) محاسبه شده است. و به ازای هر تعداد خوشه ای که $p(k)$ حداکثر گردیده است، مقدار K به عنوان تعداد خوشه در نظر گرفته شده است. در این تحقیق محدوده پیشنهادی K با توجه به نظر افراد خبره به صورت $30 \leq k \leq 50$ در نظر گرفته شده است.

$$k = \max_{s \leq k \leq t} \{P(k)\} \quad (۱۳)$$

بدین ترتیب می توان تعداد خوشه ای را انتخاب نمود که به ازای آن فاصله بین مراکز خوشه ها و شباهت مراکز خوشه با اعضای درون هر خوشه حداکثر گردد [۷].

۲-۸. بهبود تابع DEB

برای اندازه گیری میزان شباهت مشتریان / مناطق باید همه عوامل شباهت به صورت هم زمان در نظر گرفته شود. در تابع DCB دو عامل گروه کالای مشابه و ارزش حجمی در نظر گرفته شده اند. در تابع فاصله اقلیدسی عامل فاصله مکانی بین دو مشتری / منطقه در نظر گرفته شده است. این سه عامل به کمک رابطه (۱۴) ترکیب شده اند.

$$\alpha(C_i, C_j) = \frac{\text{Max}d(C_i, C_j) - d(C_i, C_j)}{\text{Max}d(C_i, C_j)} \quad (۱۴)$$

مقدار α همواره بین صفر و یک می باشد. اگر $C_i = C_j$ باشد مقدار $\alpha = 1$ است و اگر دو مشتری / منطقه اگر C_i, C_j حداکثر فاصله را داشته باشند مقدار $\alpha = 0$ است. میزان شباهت دو مشتری / منطقه اگر C_i, C_j با توجه به سه عامل فاصله مکانی، گروه کالای مشابه و میزان ارزش حجمی سفارشات به صورت رابطه (۱۵) حاصل می شود.

یکی از روش های ارزیابی عملکرد خوشه بندی، روش سنجش تراکم درون خوشه ای است. این معیار میزان تراکم خوشه ها را زمانی که تعداد خوشه ها ثابت است نشان می دهد. در این معیار باید متغیرها در یک محدوده باشند. با توجه به این که در این تحقیق داده ها گسسته هستند برای محاسبه تراکم یک خوشه، فاصله مشتریان (مناطق) درون یک خوشه باید با مرکز خوشه سنجیده شود. تراکم خوشه I_n ، I_n ، به صورت رابطه (۷) تعریف شده است:

$$I_n = \sum_{C_i \in O^n} \text{Dist}(C_i, C^n) \quad (۷)$$

سپس $F(K)$ برای K خوشه برابر است با:

$$F(K) = \frac{1}{K} \sum_{n=1}^k \sum_{C_i \in O^n} \text{Dist}(C_i, C^n) \quad (۸)$$

در واقع $F(K)$ میانگین مربعات فاصله اقلیدسی بین هر مشاهده و میانه خوشه ای که به آن تعلق دارد، است هرچه $F(K)$ کمتر باشد نشان می دهد که خوشه ها متراکم ترند و خوشه بندی بهتر صورت گرفته است [۸].

۲-۷. انتخاب بهترین تعداد خوشه

همانطور که بیشتر اشاره شد، بهترین خوشه بندی، خوشه بندی است که مجموع تشابه بین مرکز خوشه و همه شهرهای عضو خوشه مربوطه حداکثر و مجموع تشابه بین مراکز خوشه ها را حداقل نماید.

اگر $O = \{C^n | n = 1, \dots, k\}$ مجموعه مراکز خوشه ها و C^n مراکز خوشه ها باشد و $O^n = \{C_i | i = 1, \dots, |T^c - O|\}$ مجموعه باقیمانده مشتریان (مناطق) که به عنوان مرکز خوشه انتخاب نشده اند و T^c مجموعه کلیه شهرهایی که خوشه بندی روی آن ها صورت گرفته است باشد، کیفیت نتایج خوشه بندی با K خوشه می تواند به صورت معادله (۹) تعریف شود:

$$p(K) = \frac{1}{K} \sum_{n=1}^k \left(\text{Min}_{1 \leq m \leq k, m \neq n} \left\{ \frac{\mu_n + \mu_m}{\delta_{nm}} \right\} \right) \quad (۹)$$

$$\mu_n = \frac{1}{\|O^n\|} \sum_{C_i \in O^n} \text{sim}(C_i, C^n) \quad (۱۰)$$

$$\mu_m = \frac{1}{\|O^m\|} \sum_{C_j \in O^m} \text{sim}(C_j, C^m) \quad (۱۱)$$

بر مبنای این معیار، ارزش هر مشتری / منطقه را می توان به وسیله رابطه (۱۸) تعیین نمود.

$$V(C_i) = W^D \cdot D(C_i) + W^T \cdot T(C_i) + W^M \cdot M(C_i) \quad (18)$$

در معادله (۱۸) مقادیر $D(C_i)$ ، $T(C_i)$ و $M(C_i)$ به ترتیب بیانگر امتیازات مشتری / منطقه (C_i) با توجه به معیارهای D ، T و M است. W^D ، W^T و W^M اهمیت وزن ها برای معیارهای D ، T و M را به ترتیب نشان می دهد. علاوه بر $W^M + W^T + W^D = 1$ می باشد [۹]. امتیازات معمولاً به نوع کار بردها و رویکرد امتیاز دهی وابسته است. امتیازاتی را که از بانک اطلاعاتی استخراج می نمایم قبل از محاسبه ارزش هر مشتری / منطقه ابتدا نرمال می گردند. بنابر این $D(C_i)$ ، $T(C_i)$ و $M(C_i)$ را به وسیله روابط (۱۹) تعریف می کنیم.

$$\begin{aligned} D(C_i) &= \frac{Q^D - Q^D_{Min}}{Q^D_{MAX} - Q^D_{Min}} \\ T(C_i) &= \frac{Q^T - Q^T_{Min}}{Q^T_{MAX} - Q^T_{Min}} \\ M(C_i) &= \frac{Q^M - Q^M_{Min}}{Q^M_{MAX} - Q^M_{Min}} \end{aligned} \quad (19)$$

در روابط (۱۹) مقادیر Q^D و Q^T و Q^M مقدار اصلی امتیازات مشتری / منطقه (C_i) مطابق با تعریف D ، T و M هستند. مقادیر Q^D_{Min} ، Q^T_{Min} و Q^M_{Min} حداقل مقدار D ، T و M و Q^D_{Max} ، Q^T_{Max} و Q^M_{Max} حداکثر مقدار D ، T و M همه مشتریان / مناطق را نشان می دهند. سود آوری هر خوشه Q^n با محاسبه ارزش همه مشتریان / مناطق خوشه n حاصل می شود. این ارزش به صورت معادلات (۲۰ و ۲۱) محاسبه می شود:

$$V(O^n) = W^D \cdot W(O^n) + W^T \cdot T(O^n) + W^M \cdot M(O^n) \quad (20)$$

$$\begin{aligned} W(O^n) &= \frac{\sum_{C_i \in O^n} D(C_i)}{\|O^n\|} & T(O^n) &= \frac{\sum_{C_i \in O^n} T(C_i)}{\|O^n\|} \\ M(O^n) &= \frac{\sum_{C_i \in O^n} M(C_i)}{\|O^n\|} \end{aligned} \quad (21)$$

در معادلات (۲۱) مقادیر $D(O^n)$ ، $T(O^n)$ ، $M(O^n)$ بیانگر امتیازات خوشه n با توجه به معیارهای D ، T و M است. پس از مشخص شدن سود آوری هر خوشه به راین اساس مرتب می شود و

$$\left[\begin{aligned} sim(C_i, C_j) &= \frac{\sum_{a=1}^s \sum_{b=1}^t m_{ia} \times m_{jb} \times \text{Min}[\delta(g_{ia}, g_{jb}), \alpha(C_i, C_j)]}{\sum_{a=1}^s \sum_{b=1}^t m_{ia} \cdot m_{jb}} & C_i \neq C_j \\ sim(C_i, C_j) &= 1 & C_i = C_j \end{aligned} \right] \quad (15)$$

متعاقباً فاصله دو شهر / منطقه یا تفاوت رفتار آن ها به کمک رابطه (۴) محاسبه می شود. همچنین برای محاسبه شباهت کلیه داده های ۳۶ ماه گذشته با یکدیگر جمع شده اند و زمان که عامل مهمی است با مجتمع کردن داده ها عملاً نادیده گرفته شده است و وابستگی بین اقلام صادر شده در سالها و فصل های مختلف در نظر گرفته نمی شود. لذا در این مرحله برای محاسبه ماتریس فاصله، داده های هر ماه جداگانه مورد بررسی قرار گرفته و برای هر ماه ماتریس فاصله جداگانه محاسبه شده است. سپس با استفاده از روش دلفی و نظرات خبرگان ضرایب وزنی هر ماه محاسبه گردیده و ماتریس فاصله کل برای ۳۶ ماه بر حسب تابع بهبود یافته DCB محاسبه شده است. ماتریس فاصله در ماه آخر ماتریس فاصله در ماه اول

$$\left[\begin{array}{cccc} d_{11} & d_{12} & d_{1j} & d_{1n} \\ d_{21} & d_{22} & d_{2j} & d_{2n} \\ d_{i1} & d_{i2} & d_{ij} & d_{in} \\ d_{n1} & d_{n2} & d_{nj} & d_{nn} \end{array} \right] \dots \dots \left[\begin{array}{cccc} d_{11} & d_{12} & d_{1j} & d_{1n} \\ d_{21} & d_{22} & d_{2j} & d_{2n} \\ d_{i1} & d_{i2} & d_{ij} & d_{in} \\ d_{n1} & d_{n2} & d_{nj} & d_{nn} \end{array} \right] \quad (16)$$

مقدار ماتریس فاصله کل ۳۶ ماه (DIS) از فرمول (۱۷) محاسبه می شود.

$$Dis = \sum_{Y=1}^Y W_Y \cdot D_Y \quad (17)$$

D_Y ماتریس فاصله ماه y و W_Y ضریب ماه y بین $1 \leq Y \leq 36$ می باشد.

۳. مدل DTM برای ارزیابی خوشه ها

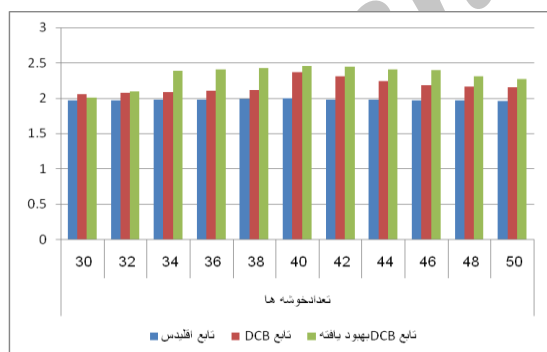
مدل DTM ارزش مشتریان / مناطق را بر مبنای سه معیار فاصله مکانی (D)، فاصله زمانی بین سفارشات (T) و ارزش حجمی (M) اندازه گیری می کند. متغیر فاصله مکانی، میانگین فاصله بین مشتریان و مرکز خوشه را اندازه می گیرد. متغیر فاصله زمانی بین سفارشات، متوسط زمانی بین سفارشات در هر خوشه را نشان می دهد. متغیر ارزش حجمی بیانگر متوسط حجم کالای درخواستی در هر خوشه می باشد. مدل DTM به سازمان ها کمک می کند تا بتوانند ارزش خوشه های خود را در این مدل مشخص نمایند و خوشه های هدف را جهت برنامه ریزی تعیین نموده و اقدامات لازم در هر خوشه را برای کاهش هزینه تعیین نمایند.

۴-۲. محاسبه تعداد خوشه‌ها

برای تعیین تعداد خوشه‌ها با استفاده از معادله (۹) مقدار $F(k)$ برای مقادیر مختلف (K) محاسبه شده است مقدار $F(k)$ حداکثر نشان دهنده تعداد مناسب خوشه‌هاست. در این تحقیق $F(k)$ برای مقادیر $30 \leq K \leq 50$ محاسبه شده است که این بازه با توجه به نظر خبرگان و تعداد داده‌ها تعیین شده است. بدین منظور به کمک نرم افزار R و ماتریس فاصله محاسبه شده برای $30 \leq K \leq 50$ خوشه بندی صورت گرفته است. کیفیت خوشه‌ها در به ازه $30 \leq K \leq 50$ در جدول (۲) نشان داده شده است.

جدول ۲. کیفیت خوشه‌ها در بازه $30 \leq K \leq 50$

تعداد خوشه	تابع اقلیدسی	تابع DCB	تابع DCB بهبود یافته
۳۰	۱/۹۷۰	۲/۰۵۱	۲/۰۱۱
۳۲	۱/۹۷۲	۲/۰۷۲	۲/۰۹۵
۳۴	۱/۹۷۸	۲/۰۸۰	۲/۳۹۰
۳۶	۱/۹۸۰	۲/۱۰۷	۲/۴۰۲
۳۸	۱/۹۸۳	۲/۱۱۱	۲/۴۳۰
۴۰	۱/۹۹۲x	۲/۳۶۹x	۲/۴۵۲x
۴۲	۱/۹۸۱	۲/۳۱۰	۲/۴۴۰
۴۴	۱/۹۷۵	۲/۳۴۲	۲/۴۰۱
۴۶	۱/۹۷۲	۲/۱۸۲	۲/۳۹۵
۴۸	۱/۹۶۵	۲/۱۶۱	۲/۳۱۲
۵۰	۱/۹۶۰	۲/۱۵۴	۲/۲۷۰



نمودار ۱. مقایسه کیفیت و خوشه‌ها

۴-۳. تحلیل خوشه‌ها در مدل DTM

برای تحلیل خوشه‌ها در مدل DTM باید معیارهای M, T, D برای مشتریان / مناطق و خوشه‌ها محاسبه گردد. مقدار متغیر فاصله اقلیدسی بین مشتریان / مناطق نشان دهنده فاصله اقلیدسی است که بین دو مشتری / منطقه وجود دارد. این متغیر برای هر

$$D(C_i) = \frac{Q^D - Q_{Min}^D}{Q_{MAX}^D - Q_{Min}^D}$$

با ارزش‌ترین آن‌ها مشخص شده و به شرکت در برنامه ریزی و تعیین استراتژی توزیع هر خوشه و خدمت رسانی به مشتریان هدف شرکت کمک می‌نماید.

۴. مطالعه موردی

مورد مطالعاتی در این تحقیق شرکت ایساکو می‌باشد. شرکت ایساکو در راستای برنامه‌های کلان گروه صنعتی ایران خودرو در پی توسعه و تقویت سیستم حمل و نقل قطعات یدکی و دستیابی به سهم بیشتر بازار هدف به شکل اثر بخش می‌باشد. این شرکت با بکارگیری دانش و فناوری جدید به دنبال ایجاد بستر مناسب جهت تقویت زیربنای لازم برای تسهیل و توسعه سیستم حمل و نقل شرکت است. مطالعه صورت گرفته بر روی داده‌های ایساکو به مشتریان / مناطق شهر تهران در طی ۳۶ ماه گذشته است. این شرکت طیف وسیعی از کالاهای متنوع را بین مشتریان / مناطق توزیع می‌کند. تعداد اقلام مورد بررسی ۷۵ قلم کالای مهم بوده که در ۸ گروه دسته بندی شده‌اند. داده‌های شرکت ایساکو در طی سه سال گذشته با توجه به ۷۰۰ نمایندگی و ۸۰۰ فروشگاه شامل $(50)(800)=75000$ رکورد می‌باشد. هر مشتری حداقل هفته‌ای یک سفارش دارد و هر سال معادل (۵۰) هفته فرض شده است.

۴-۱. خوشه بندی مناطق / مشتریان

بعد از انجام مراحل آماده سازی، یکپارچگی و استخراج داده‌ها با فرمت مناسب با استفاده از الگوریتم k -means مناطق کشور خوشه بندی شده‌اند. همانطور که توضیح داده شد خوشه بندی یک بار بر اساس فاصله اقلیدسی، یک بار بر اساس تابع DCB و یکبار بر اساس تابع DCB بهبود یافته صورت گرفته است. تعداد خوشه‌ها برابر ۴۰ در نظر گرفته شده است که در ادامه علت این انتخاب توضیح داده شده است. تراکم خوشه‌های حاصل شده در سه روش با توجه به معادله (۸) محاسبه شده است. هرچه مقدار $F(K)$ در معادله (۸) کمتر باشد میزان تراکم خوشه‌ها بیشتر است. جدول (۱) میزان تراکم خوشه‌ها با استفاده از سه روش پیشنهادی را نشان می‌دهد. در جدول زیر مشاهده می‌شود که تراکم خوشه‌ها در روش تابع DCB بهبود بیشتر از دو تابع دیگر است. یعنی این تابع نسبت به توابع دیگر عملکرد بهتری داشته است.

جدول ۱. مقایسه میزان تراکم خوشه‌ها از توابع مختلف

تراکم خوشه	تابع فاصله خوشه بندی
۱/۲۹۴	تابع فاصله اقلیدسی
۰/۱۱۹	تابع DCB
۰/۰۷۴	تابع DCB بهبود یافته

به اهمیت این متغیرها به را ی متغیرهای M,T,D به ترتیب وزن های ۰/۵، ۰/۳، ۰/۲ در نظر گرفته شده است. سپس با معادلات $M(O^n)$ ، $T(O^n)$ ، $D(O^n)$ مقدار فاصله مکانی، فاصله زمانی و ارزش حجمی هر خوشه محاسبه و با استفاده از رابطه $V(O^n)$ ارزش هر خوشه تعیین می گردد. نتایج حاصل به صورت جدول (۳) آمده است:

محاسبه می شود. متغیر فاصله زمانی بین سفارشات مشتریان / مناطق نشان دهنده فاصله بین دو سفارش متوالی مشتریان / مناطق است. مقدار این متغیر برای هر مشتری / منطقه با توجه به رابطه $T(C_i)$ محاسبه می شود. متغیر ارزش حجمی کالاهای درخواستی نیز بیانگر میزان حجم کالای درخواستی توسط مشتری / منطقه می باشد که توسط رابطه $M(C_i)$ محاسبه می شود. وزن متغیرها بر اساس نظرات خبرگان تعیین گردیده است. با توجه

جدول ۳. ارزش مشتریان / مناطق هر خوشه

شماره خوشه	میانگین فاصله مکانی $D(O^n)$	میانگین فاصله زمانی $T(O^n)$	میانگین ارزش حجمی $M(O^n)$	میانگین ارزش خوشه	تعداد مشتری / منطقه در خوشه
۱۷	۰/۴۱۷	۰/۷۲۱	۰/۵۴۰	۰/۵۳۲۸	۱۵
۲۹	۰/۹۱۸	۰/۱۵۱	۰/۲۷۱	۰/۵۵۸۵	۱۸
۸	۰/۸۴۱	۰/۲۷۴	۰/۲۵۲	۰/۵۵۳۱	۲۵
۳۰	۰/۵۷۱	۰/۱۷۱	۰/۴۷۱	۰/۴۳۱	۴۲
۱۸	۰/۶۴۱	۰/۲۴۸	۰/۵۴۲	۰/۵۰۳۳	۱۱
۳۱	۰/۵۱۷	۰/۲۱۵	۰/۷۲۱	۰/۴۶۷۲	۵
۷	۰/۶۱۸	۰/۲۱۸	۰/۵۱۱	۰/۴۷۶۶	۲
۲۷	۰/۴۲۵	۰/۴۱۸	۰/۹۲۱	۰/۵۲۲۱	۷
۹	۰/۴۱۹	۰/۴۷۱	۰/۴۲۱	۰/۴۳۵	۴۸
۳۲	۰/۲۷۱	۰/۵۷۱	۰/۲۰۹	۰/۳۴۸۶	۱۵
۱۹	۰/۲۱۵	۰/۲۴۸	۰/۶۲۱	۰/۳۰۶۱	۱۸
۴	۰/۴۹۵	۰/۵۷۱	۰/۵۲۱	۰/۵۲۳	۱۹
۲۶	۰/۷۱۵	۰/۶۹۱	۰/۷۴۱	۰/۷۱۳	۲۱
۱۰	۰/۸۲۱	۰/۵۷۱	۰/۸۱۱	۰/۷۴۴	۲۲
۳۳	۰/۹۴۲	۰/۸۲۱	۰/۹۱۵	۰/۹۰۰۳	۱۷
۳	۰/۹۱۲	۰/۲۸۱	۰/۹۴۷	۰/۷۲۹۷	۱۶
۳۴	۰/۱۵۰	۰/۲۱۵	۰/۰۱۷	۰/۱۴۲۹	۱۹
۱۱	۰/۱۲۷	۰/۲۵۲	۰/۲۱۵	۰/۱۸۲۱	۱۱
۳۵	۰/۹۱۶	۰/۹۴۷	۰/۵۶۱	۰/۸۵۴۳	۱۷
۲	۰/۱۹۷	۰/۱۵۷	۰/۲۴۸	۰/۱۹۵۲	۲۵
۲۰	۰/۸۱۷	۰/۸۱۵	۰/۱۱۱	۰/۶۷۵۲	۴۲
۱	۰/۲۴۸	۰/۲۴۹	۰/۱۵۲	۰/۲۲۹۱	۵۱
۲۱	۰/۲۵۲	۰/۰۵۰	۰/۱۱۷	۰/۱۶۴۴	۷۱
۱۲	۰/۵۴۲	۰/۰۴۸	۰/۹۴۸	۰/۴۷۵	۱۵
۴۰	۰/۷۴۲	۰/۰۲۵	۰/۵۲۱	۰/۴۸۲۷	۶۸
۲۲	۰/۵۴۱	۰/۲۱۵	۰/۷۱۸	۰/۴۷۸۶	۱۵
۵	۰/۸۲۱	۰/۱۸۷	۰/۱۵۱	۰/۴۹۶۸	۱۷
۲۴	۰/۲۷۱	۰/۹۴۱	۰/۱۷۲	۰/۴۵۲۲	۱۶
۳۹	۰/۲۵۷	۰/۱۵۱	۰/۱۵۱	۰/۲۰۴	۵
۱۴	۰/۴۲۲	۰/۲۴۸	۰/۹۴۱	۰/۴۷۳۶	۷
۳۶	۰/۴۴۷	۰/۶۴۷	۰/۱۵۷	۰/۴۴۹	۹
۶	۰/۵۷۴	۰/۱۴۸	۰/۲۴۸	۰/۳۸۱	۸
۳۷	۰/۱۱۵	۰/۰۰۵	۰/۷۴۸	۰/۲۰۸۶	۴
۲۳	۰/۵۴۲	۰/۱۵۶	۰/۱۵۱	۰/۳۴۸	۳
۱۶	۰/۶۷۱	۰/۱۷۸	۰/۱۷۵	۰/۴۲۳۹	۱۵
۲۵	۰/۱۲۵	۰/۲۴۹	۰/۱۶۱	۰/۱۶۹۴	۱۷
۳۸	۰/۷۶۴	۰/۳۴۸	۰/۷۴۰	۰/۶۳۴۴	۱۴
۱۳	۰/۱۱۷	۰/۷۱۵	۰/۷۴۵	۰/۴۲۲	۲۵
۲۸	۰/۹۱۵	۰/۸۴۸	۰/۲۱۱	۰/۷۵۴۱	۱۶
۱۵	۰/۸۲۱	۰/۲۴۱	۰/۹۱۱	۰/۶۶۵	۱۹

استراتژی های توزیع در هر خوشه میانگین ارزش حجمی هر گروه کالا را به صورت جداگانه محاسبه و سپس بر میانگین ارزش

پس از تحلیل انجام شده بر روی خوشه ها و محاسبه ارزش هر خوشه استراتژی های توزیع استخراج می گردد. برای تعیین

بهره گرفت. همچنین می‌توان تحقیقات را بر روی گروه کالای خاص با شرایط مختلف انجام داد و نتایج را بر روی هر یک از گروه محصولات خاص بررسی نمود.

حجمی کل تقسیم می‌کنیم. در اینصورت ارزش حجمی هر گروه کالا نسبت به کل اقلام به کمک رابطه (۲۲) محاسبه می‌شود. میانگین ارزش گروه کالای a برای هر خوشه

میانگین ارزش برای هر خوشه

$$W_a = \frac{\text{ارزش گروه } a}{\text{ارزش کل}} \quad (22)$$

$$W_{a1} + W_{a2} + \dots + W_{av} = 1$$

بر اساس مقادیر W_a ظرفیت حجمی وسایل توزیع بین گروه کالاهای مختلف در هر خوشه به صورت مناسب تقسیم می‌شود. محدودیت ظرفیت وسایل نقلیه یک محدودیت مشترک بین همه گروه کالاهای مورد برنامه ریزی می‌باشد. این نسبت‌ها بعد از هر بار اجرای برنامه با توجه به اطلاعات جدید به هنگام می‌شود. بدین منظور ابتدا ظرفیت باقیمانده وسایل نقلیه برای هر گروه کالا در هر خوشه محاسبه می‌شود. سپس از جمع این ظرفیت‌های باقیمانده برای هر گروه کالا کل ظرفیت باقیمانده محاسبه می‌شود. ظرفیت باقیمانده برای گروه کالاهایی که زیاد محاسبه شده کم می‌شود و به گروه کالاهای که ظرفیت باقیمانده کمتری دارند تخصیص داده می‌شود. بدین صورت ظرفیت باقیمانده بین گروه‌های مختلف کالا تسطیح و از ظرفیت وسایل نقلیه حداکثر استفاده به عمل می‌آید.

۵. نتیجه گیری

با افزایش اهمیت رضایت مشتری در محیط تجاری امروز، بسیاری از سازمان‌ها بر روی مباحث مرتبط با شناخت مشتری، وفاداری و سودآوری مشتری برای افزایش سهم بازار خود و کسب رضایت مشتری تمرکز نموده‌اند. مدیریت ارتباط با مشتری به عنوان یک مزیت رقابتی برای سازمان‌ها محسوب می‌گردد. یکی از روش‌های شناخت مشتریان، بخش بندی مشتریان به گروه‌های همگن و اتخاذ سیاست‌های بازاریابی مناسب با هر بخش است. در این مقاله یک تابع فاصله رفتار مشتریان (DCB) بر اساس ترکیب قواعد تعریف گردیده است. و از آن به عنوان تابع سنجش فاصله در الگوریتم خوشه بندی K-means استفاده شده است که نتایج را نسبت به تابع فاصله اقلیدسی به میزان قابل توجهی بهبود داده است. خوشه بندی صورت گرفته با تابع سنجش تراکم با یکدیگر مقایسه شده‌اند و تعداد خوشه‌ها با استفاده از تابع سنجش کیفیت مشخص شده است. پس از بخش بندی مشتریان / مناطق با استفاده از مدل DTM ارزش هر خوشه تعیین و سیاست‌های متناسب با هر بخش تعیین گردیده است. در تحقیقات آتی می‌توان از داده‌ای توصیفی و جمعیت شناختی مشتریان / مناطق نیز در تعریف و بهبود تابع DTM

منابع

- [1] Alex Berson, Stephen Smith, kurt Thearling, "Bulding Data Mining Application for CRM", McGraw-Hill, 2001.
- [2] Nong, Y., "The Handbook of Data Minig", LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS Mahwah, New Jersey London, 2003.
- [3] Hsiao-Fan Wang, Wei-Kuo Hong, 1. "Managing Customer Profitability in a Competitive Market by Continuous Data Mining", Department of Industrial Engineering and Engineering Management. National Tsing Hua University, Hsinchu, Taiwan, ROC, 2005.
- [4] Chris Rygielski, a., Jyun-Cheng, Wang, b., David, C., Yen, a., "Data Mining Techniques for Customer Relationship Management", Technology in Society 24, 2002, pp. 483-502
- [5] Hyunseok Hwang, Taesoo Jung, Euiho Suh, "An LTV Model and Customer Segmentation Based on Customer Value: a Case Study on the Wireless Telecommunication Industry", Expert Systems with Applications 26, 2004, pp. 181-188.
- [6] Su-Yeon kim, Tae-see Jung, Eui-Ho Suh, Hyun-seok Hwang, "Customer Segmentation and Strategy Development Based on Customer Lifetime Value: A Case Study y", Expert Systems with Applications 31, 2006, pp. 101-107
- [7] Tsai, C.Y., chiu, C.C., "A Purchase - Based Market Segmentation Methodology", Expert Systems with Applications 27, 2004, pp. 265-276.
- [8] Shina, H.W., Sohn, S.Y., "Segmentation of Stock Trading Customers According to Potential Value.", Expert Systems with Applications 27, 2004, pp. 27-33.
- [9] Jedid-Jah Jonkera, Nanda Piersmab, Dirk Van den Poelc, "Joint Optimization of Customer Segmentation and Marketing Policy to Maximize Long-Term Profitability", Expert Systems with Applications 27, 2004, pp. 159-168.
- [10] Pauline, A., Wilcox, Calin Gurau, "Business Modeling with UML: the Impiementation of CRM System for Online Retailing", Journal of Retailing and Consumer Services 10, 2003, pp. 181-191.
- [11] Wegner, A., kamakura, Michel Wedel, Fernando de Rosa, Jose Afonso Mazzon, "Cross-Selling Through Database Marketing: a to Mixed Data Factor Analyzer for Data Augmentation and Prediction", International journal of Research in Marketing 20, 2003, pp. 45-65.

- [12] Arindam Banerjee, joydeep Ghosh, "*Clickstream Clustering using Weighted Longest Common Subsequences*" Dep of Electrical Engineering university of Texas at Austin, 2002.
- [13] Jiawei Han, Micheline Kamber, Department of computer Science, University of Illinois at Urbana – champaign, [www.cs. Uiuc .edu/~hanj](http://www.cs.Uiuc.edu/~hanj).

Archive of SID