



دانشکده‌ی مهندسی صنایع و مدیریت

گروه : مدیریت

پایان نامه کارشناسی ارشد

خوشه‌بندی فازی مشتریان و تحلیل رفتار آن‌ها با رویکرد داده‌کاوی پویا

(مطالعه موردی : تلفن همراه سامسونگ)

دانشجو : مریم علائی‌زاده

استاد راهنما :

دکتر رضا شیخ

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

بهمن ۱۳۹۳

بسمه تعالی

شماره :

تاریخ :

فرم صورتجلسه دفاع از پایان نامه تحصیلی دوره کارشناسی ارشد

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عج) ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد خانم مریم علایی زاده رشته MBA گرایش عمومی. تحت عنوان خوشه بندی فازی مشتریان و تحلیل رفتار آن ها با رویکرد داده کاوی پویا که در تاریخ ۱۳۹۳/۱۱/۲۸ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می گردد:

<input type="checkbox"/> قبول (با درجه : عالی امتیاز : ۱۹)	<input type="checkbox"/> دفاع مجدد	<input type="checkbox"/> مردود
--	------------------------------------	--------------------------------

۱- عالی (۲۰ - ۱۹)

۲- بسیار خوب (۱۸/۹۹ - ۱۸)

۳- خوب (۱۷/۹۹ - ۱۶)

۴- قابل قبول (۱۵/۹۹ - ۱۴)

۵- نمره کمتر از ۱۴ غیر قابل قبول

عضو هیأت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
۱- استاد راهنما	دکتر رضا شیخ	استادیار	
۲- استاد مشاور	مهندس طحانیان		
۳- نماینده شورای تحصیلات تکمیلی	دکتر سعید حکمی نسب		
۴- استاد ممتحن	دکتر میرلوحی		
۵- استاد ممتحن	دکتر دهقانی		

رئیس دانشکده

مشکرو قدردانی

با سپاس از وجود مقدس آمان که

ناتوان شدند تا ما به توانایی برسیم...

موبایشان سپید شد تا ما رو سفید شویم...

و عاشقانه سوختند تا گرما بخش وجود ما رو و سگر راهمان باشند...

پدر عزیزم

مادر مهربانم

همسر بردبارم

استادانم که علم و معرفت را به من آموختند.

مریم علانی زاده

بهمن ۹۳

تعهدنامه

اینجانب مریم علائی زاده دانشجوی دوره کارشناسی ارشد رشته MBA دانشکده مهندسی صنایع و مدیریت دانشگاه شاهرود نویسنده پایان نامه "خوشه بندی فازی مشتریان و تحلیل رفتار آن ها با رویکرد داده کاوی پویا (مطالعه موردی : تلفن همراه سامسونگ)" تحت راهنمایی جناب آقای دکتر رضا شیخ متعهد می شوم:

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از اصالت برخوردار است.
- در استفاده از نتایج پژوهش های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان نامه تا کنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه شاهرود می باشد و مقالات مستخرج با نام دانشگاه شاهرود و یا "Shahrood University" به چاپ خواهد رسید.
- حقوق معنوی تمامی افرادی که در به دست آوردن نتایج اصلی پایان نامه تاثیرگذار بوده اند، در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام پایان نامه، در مواردی که از موجود زنده (یا بافت های آن) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.

در کلیه مراحل انجام این پایان نامه در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته و یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاقی رعایت شده است.

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزارها و تجهیزاتی که ساخته شده است) متعلق به دانشگاه شاهرود می باشد. ایم مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده فارسی

امروزه مشتری به عنوان کلید اصلی موفقیت یا عدم موفقیت یک شرکت محسوب می‌شود. لذا مطالعه رفتار مشتریان طی دهه‌های اخیر بسیار مورد توجه محققان بازاریابی قرار گرفته است. از مهم‌ترین زمینه‌های مطالعه در این حوزه، بررسی وفاداری، رضایت، ترجیحات و سلیقه مشتریان می‌باشد که محققان تلاش می‌کنند تا با بکارگیری روش‌ها و ابزارهای مختلف رفتار مشتری را مورد تحلیل قرار دهند. این پژوهش با توجه به هدف کاربردی و از نظر روش شناسی جز تحقیق‌های پیمایشی و از نظر موضوعی جز تحقیقات بازاریابی محسوب می‌شود. هدف اصلی این تحقیق، بخش بندی مشتریان با استفاده از تکنیک‌های داده‌کاوی و تحلیل و پیش بینی رفتار آن‌ها با استفاده از الگوریتم C-means پویا می‌باشد. برای این منظور داده‌های خرید ۳۰۰۰ نفر از کاربران گوشی همراه سامسونگ از پایگاه داده شرکت دریافت شده و پس از کد نویسی الگوریتم C-means با استفاده از نرم افزار Matlab 2008a مورد بررسی قرار گرفته است. نتایج نشان می‌دهد که به کارگیری روش پویا بخش بندی واقعی تری نسبت به روش ایستا ارائه می‌دهد همچنین تحلیل رفتار در فرایند زمانی نشان می‌دهد که مشتریان نسبت به قیمت گوشی حساس بوده و ویژگی‌های کیفیت دوربین و پردازشگر و تعداد سیم کارت و RAM علل تغییر در بخش بندی مشتریان می‌باشند.

واژه‌های کلیدی: رفتار مشتری، داده‌کاوی پویا، خوشه بندی، FCM

لیست مقالات استخراج شده از پایان نامه:

رویکردی نوین در تصمیم سازی بازاریابی با تمرکز بر رفتارهای متغیر مشتریان در شرایط عدم

اطمینان

فهرست مطالب

عنوان	صفحه
فصل اول: کلیات تحقیق	۱
۱-۱- مقدمه	۳
۲-۱- بیان مسئله	۴
۳-۱- اهمیت و ضرورت تحقیق	۷
۴-۱- هدف تحقیق	۷
۵-۱- سوالات تحقیق	۸
۶-۱- روش تحقیق	۸
۷-۱- قلمرو تحقیق	۸
۱-۷-۱- قلمرو موضوعی	۸
۲-۷-۱- قلمرو مکانی	۹
۳-۷-۱- قلمرو زمانی	۹
۸-۱- تعریف واژه ها و اصلاحات	۹
۱-۸-۱- داده کاوی	۹
۲-۸-۱- خوشه بندی	۹
۳-۸-۱- داده کاوی پویا	۹
۴-۸-۱- منطق فازی	۱۰
فصل دوم: ادبیات تحقیق	۱۱
۱-۲- مقدمه	۱۳
۲-۲- داده کاوی	۱۳
۳-۲- خوشه بندی	۱۶
۱-۳-۲- هدف خوشه بندی	۱۸
۲-۳-۲- انواع خوشه بندی	۱۸
۳-۳-۲- الگوریتم های خوشه بندی	۲۲
۴-۳-۲- معیار فاصله	۲۳
۵-۳-۲- الگوریتم C-means	۲۴
۶-۳-۲- الگوریتم FCM	۲۵
۴-۲- طبقه بندی	۲۹

۳۰	۵-۲- کشف قواعد وابستگی
۳۰	۶-۲- کشف الگوهای ترتیبی
۳۱	۷-۲- داده کاوی پویا
۳۲	۸-۲- خوشه بندی پویا
۳۴	۹-۲- روش های نوین برای خوشه بندی فازی پویا
۳۶	۱-۹-۲- اختصاص اعضای پویا به کلاس های ایستا
۳۶	۱-۱-۹-۲- روش به کار بردن مسیر در خوشه بندی
۳۷	۲-۱-۹-۲- تشریح عملکرد C- میانگین فازی
۳۹	۲-۹-۲- اختصاص اعضای ایستا به کلاس پویا
۴۱	۱-۲-۹-۲- شرح جزئیات روش پیشنهاد شده
۴۸	۱۰-۲- منطق فازی
۴۸	۱-۱۰-۲- مجموعه های فازی
۵۱	۲-۱۰-۲- میانگین فازی
۵۱	۱-۲-۱۰-۲- میانگین مثلثی
۵۲	۲-۲-۱۰-۲- میانگین دوزنقه ای
۵۲	۳-۱۰-۲- فازی زدایی از میانگین فازی
۵۳	۱۱-۲- بخش بندی و تحلیل رفتار مشتریان
۵۷	فصل سوم: روش تحقیق
۵۹	۱-۳- مقدمه
۵۹	۲-۳- نوع پژوهش
۶۰	۳-۳- روش گردآوری داده ها
۶۱	۴-۳- جامعه و نمونه آماری
۶۳	۵-۳- روش تجزیه و تحلیل داده ها
۶۹	فصل چهارم: تجزیه و تحلیل داده ها
۷۱	۱-۴- مقدمه
۷۱	۲-۴- تجزیه و تحلیل دوره ای داده ها
۷۲	۱-۲-۴- تحلیل داده های دوره اول
۷۴	۲-۲-۴- تحلیل داده های دوره اول و دوم
۷۶	۳-۲-۴- تحلیل داده های دوره اول و دوم و سوم
۷۸	۴-۲-۴- تحلیل داده های چهار دوره
۸۰	۵-۲-۴- تحلیل داده های پنج دوره

۸۲ تجزیه و تحلیل تجمعی داده ها
۸۲ ۱-۳-۴ تحلیل قطعی
۸۳ ۲-۳-۴ تحلیل فازی
۸۴ ۳-۵ تجزیه و تحلیل تغییرات رفتاری
۸۷ فصل پنجم : نتیجه گیری و پیشنهادات
۸۹ ۱-۵ مقدمه
۸۹ ۲-۵ بحث و نتیجه گیری
۹۰ ۳-۵ محدودیت های تحقیق
۹۱ ۴-۵ پیشنهادات کاربردی - مدیریتی
۹۲ ۵-۵ پیشنهاد تحقیقات آتی
۹۳ پیوست
۱۰۴ منابع

فهرست اشکال

عنوان	صفحه
شکل (۱-۲) روش های داده کاوی	۱۶
شکل (۲-۲) خوشه بندی وسایل نقلیه	۱۷
شکل (۳-۲) انواع الگوریتم های خوشه بندی	۲۲
شکل (۴-۲) معیارهای تشابه براساس توابع فاصله مختلف	۲۴
شکل (۵-۲) توزیع یک بعدی نمونه ها	۲۷
شکل (۶-۲) خوشه بندی کلاسیک نمونه های ورودی	۲۸
شکل (۷-۲) خوشه بندی فازی نمونه	۲۸
شکل (۸-۲) تشریح عملکرد FFCM	۳۹
شکل (۹-۲) نمای کلی روش خوشه بندی فازی پویا با تغییر ترکیب کلاس	۴۱
شکل (۱۰-۲) عدد فازی مثلثی	۵۰
شکل (۱-۴) خوشه بندی پویا	۷۳
شکل (۲-۴) خوشه بندی ایستا	۷۳
شکل (۳-۴) خوشه بندی پویا چرخه ۱	۷۵
شکل (۴-۴) خوشه بندی ایستا چرخه ۱	۷۵
شکل (۵-۴) خوشه بندی پویا چرخه ۲	۷۷
شکل (۶-۴) خوشه بندی ایستا چرخه ۲	۷۷
شکل (۷-۴) خوشه بندی پویا چرخه ۳	۷۹
شکل (۸-۴) خوشه بندی ایستا چرخه ۳	۷۹
شکل (۹-۴) خوشه بندی پویا چرخه ۴	۸۱
شکل (۱۰-۴) خوشه بندی ایستا چرخه ۴	۸۱

فهرست جداول

عنوان	صفحه
جدول (۱-۲) روش های خوشه بندی برای داده کاوی پویا	۳۶
جدول (۱-۳) ویژگی های مشتریان و محصولات	۶۲
جدول (۲-۳) پایگاه داده	۶۳
جدول (۱-۴) مراکز خوشه های پویا برای $C=3$	۷۲
جدول (۲-۴) مراکز خوشه های ایستا برای $C=3$	۷۲
جدول (۳-۴) مجموع خطا	۷۳
جدول (۴-۴) مراکز خوشه های پویا چرخه ۱	۷۴
جدول (۵-۴) مراکز خوشه های ایستا چرخه ۱	۷۴
جدول (۶-۴) مجموع خطای چرخه ۱	۷۵
جدول (۷-۴) مراکز خوشه های پویا چرخه ۲	۷۶
جدول (۸-۴) مراکز خوشه های ایستا چرخه ۲	۷۶
جدول (۹-۴) مجموع خطای چرخه ۲	۷۷
جدول (۱۰-۴) مراکز خوشه های پویا چرخه ۳	۷۸
جدول (۱۱-۴) مراکز خوشه های ایستا چرخه ۳	۷۸
جدول (۱۲-۴) مجموع خطای چرخه ۳	۷۹
جدول (۱۳-۴) مراکز خوشه های پویا چرخه ۴	۸۰
جدول (۱۴-۴) مراکز خوشه های ایستا چرخه ۴	۸۰
جدول (۱۵-۴) مجموع خطای چرخه ۴	۸۱
جدول (۱۶-۴) مرکز خوشه های قطعی	۸۲
جدول (۱۷-۴) مرکز خوشه های فازی	۸۳
جدول (۱۸-۴) نتایج تحلیل رفتار مشتریان	۸۴

فصل اول

کلیات

۱.۱ مقدمه

مطالعه رفتار مشتریان طی دهه های اخیر یکی از مهم ترین موضوعات مورد توجه محققان بازاریابی بوده است. تحولات پیچیده و پرشتاب جهانی در عرصه های علمی و فناوری بسیاری از سازمان های موفق جهانی را بر آن داشته است تا اهداف و روش های خود را در جهت شناخت هر چه بیشتر مشتریان و تامین نیازهای آنان هدایت کنند .

امروزه مشتری به عنوان کلید اصلی موفقیت یا عدم موفقیت یک شرکت محسوب می شود. لذا درک رفتار مشتری از اهمیت زیادی برخوردار است . رفتار مصرف کننده یک فاکتور جامع در رکود و یا به جریان افتادن تمامی فعالیت های بازرگانی محسوب می گردد. نکته کلیدی موفقیت استراتژی بازاریابی هم از جنبه علمی، محلی و جهانی، درک رفتار مصرف کننده و مشتری است. این امر برای شرکت های تجاری و سازمان های غیر انتفاعی و همین طور سازمان های دولتی که قوانین مربوط به بازار را تدوین می کنند، کاربرد دارد.

گوناگونی و تنوع رفتار مشتریان به دلیل گوناگونی عوامل اثر گذار بر رفتار فرد و انگیزه فرد برای خرید می باشد. وجود گروه های متفاوت مشتری برای بازارهای یک محصول بیانگر تفاوت های گسترده میان افراد است.

رفتار خرید به هیچ وجه ساده نیست اما درک این رفتار نیز برای مدیران بسیار حیاتی است. رفتار خرید مصرف کننده (مشتری)^۱ عبارت است از رفتار خرید مصرف کنندگان نهایی، یعنی افراد و خانوارهایی که کالا و خدمات را برای اهداف مشخصی خریداری می نمایند. رفتار به معنای شناخت فرصت ها و تهدیدهای بیرونی و بهره برداری از آنها و همچنین شناخت ضعف ها و قوت های درونی در مقابل رفتار مشتری در جهت تحقق اهداف سازمان می باشد.

¹ . Consumer buying behavior

برقراری ارتباط بهتر و موثر با مشتری ناشی از شناخت فرآیند رفتار مشتری، فرآیند تصمیم‌گیری و عوامل تاثیرگذار در آن و همچنین شناسایی فرصت‌ها و تهدیدات و مشخص نمودن ضعف‌ها و قوت‌های داخلی در مقابل رفتار مشتری است.

در شکل‌گیری رفتار مشتریان و بروز تغییرات در رفتار آن‌ها عوامل مختلفی تاثیرگذار می‌باشند که از جمله مهم‌ترین آن‌ها می‌توان به عوامل دموگرافی (مشخصات مصرف‌کننده)، عوامل اجتماعی عوامل روانشناسی، عوامل اقتصادی (اسماعیل پور و آشتیانی، ۱۳۸۱) اشاره کرد. اما باید به اینکه توجه داشت که همه این عوامل پویا هستند و به طور مستمر در طول زمان تغییر می‌کنند. از این رو تغییر رفتار مشتریان نیز جنبه پویا به خود می‌گیرد. در فضای رقابتی امروزی شرکت‌ها و سازمان‌هایی موفق‌تر خواهند بود که اهداف، روش‌ها و عملکرد خود را بر اساس شناخت رفتار مشتریان و تحلیل تغییرات آن‌ها تنظیم کنند.

۲.۱ بیان مسئله تحقیق

امروزه بسیاری از شرکت‌ها و سازمان‌ها مفاهیم جدید بازاریابی را پذیرفته‌اند و طبق آن عمل می‌کنند. آنها متوجه شده‌اند که تمرکز بر نیازهای مشتریان از مفروضات اصلی گرایش بازاریابی است. بدین ترتیب مطالعه و کشف نیازهای مشتریان و تجزیه و تحلیل رفتار آن‌ها جز وظایف اصلی واحد بازاریابی است، که در نتیجه آن سلیقه بازار هدف که از لحاظ پارامترهای سن، درآمد، ذائقه، سطح تحصیلات و ... با یکدیگر متفاوت‌اند، شناسایی و کالا یا خدمت مناسب به آن بازار عرضه می‌شود.

در خصوص شناسایی و تحلیل رفتار مشتریان یکی از مهم‌ترین اقدامات شرکت‌ها، بخش بندی مشتریان^۲ به خوشه‌های رفتاری مختلف است که بر اساس آن سه نوع اطلاعات فراهم می‌شود:

^۲.Customer Segmentation

- جهت‌گیری مشتریان
- واقعیت‌های مربوط به رفتار انسانی
- جایگاه و موقعیت هر محصول در بازار هدف

به منظور شناسایی و تحلیل رفتار مشتریان و استخراج الگوهای رفتاری آن‌ها از الگوریتم‌های داده‌کاوی^۳ استفاده می‌شود. مسئله اصلی این است که چگونه می‌توان با تکنیک‌های داده‌کاوی به شناسایی و تحلیل رفتار مشتریان پرداخت. در داده‌کاوی فرض بر این است که مجموعه داده‌هایی که قرار است هر یک از اعمال یا روش‌های داده‌کاوی بر روی آنها اعمال گردد، بصورت ایستا^۴ در داخل یک پایگاه یا در یک فایل وجود دارند و در حین عمل داده‌کاوی نیز این داده‌ها تغییری نمی‌کنند. این گونه فرض‌ها انجام عمل داده‌کاوی را بسیار ساده‌تر می‌نماید. اما نکته‌ای که باید در اینجا مورد توجه قرار گیرد این مطلب است که هر چند رفتار مشتریان بر اساس نقطه زمانی ثابت می‌شود اما در فرایند دوره‌های زمانی این رفتار متغیر است به عبارت دیگر رفتار مشتری، گروه‌های مشتریان هدف و جامعه در کل، دائماً در حال تغییر می‌باشند. لذا برای داده‌کاوی بر روی داده‌هایی از این دست، نمی‌توان روش‌های سنتی و قدیمی داده‌کاوی را بکار برد بلکه این نوع داده‌ها و مجموعه داده‌ها نیاز به الگوریتم‌ها و روش‌های داده‌کاوی پویایی^۵ دارند که از پس ویژگی‌های این داده‌ها برآیند. بنابراین نیاز به نوعی داده‌کاوی پویا احساس می‌شود.

به دنبال حذف فرض ایستا بودن داده‌ها در پایگاه داده، جهت بخش بندی مشتریان به خوشه‌های رفتاری، الگوریتم خوشه بندی فازی^۶ در داده‌کاوی پویا توسط ریچارد وبر^۷ در سال ۲۰۰۷ معرفی شد.

^۳.Data mining

^۴.Static

^۵.Dynamic data mining

^۶.Fuzzy Clustering

^۷.Weber

تکنیک خوشه بندی فازی با رویکرد پویا با شناسایی رفتار مشتریان و تغییر رفتار آن ها و تحلیل روند تغییرات محیطی و ذائقه ای جامعه به مدیران و بازاریابان جهت برنامه ریزی و طراحی و توسعه محصولات یاری می رساند.

با توجه به اینکه این تکنیک با رویکرد پویا رفتار جامعه را مورد بررسی قرار می دهد به سادگی می توان وضعیت کنونی را با شرایط قبلی جامعه مورد مقایسه قرار داده و تغییرات را مرحله به مرحله مشاهده نمود و بر اساس آن روند تغییرات در آینده پیش بینی شود. این در حالی است که تا کنون برای بررسی رفتار جامعه تکنیک های داده کاوی ایستا مورد استفاده قرار می گرفتند که نتایج حاصل از آن تنها یک وضعیت ثابت را نشان داده و قابلیت مقایسه و تحلیل تغییرات را نداشته اند.

از آن جا که ماهیت جامعه پویا و دائما در حال تغییر می باشد، بخش بندی مشتریان با رویکرد ایستا نمی تواند اطلاعات مناسبی را جهت تصمیم گیری در اختیار مدیران قرار دهد، زیرا در گذر زمان شرایط مشتریان و ذائقه افراد هم دستخوش تغییر خواهند شد و بخش بندی قبلی کارایی خود را از دست خواهد داد مگر اینکه شرکت ها قادر باشند با انجام تحقیقات مستمر تغییرات را پیش بینی نمایند که البته این روش نیز به دلیل صرف هزینه و زمان بسیار، منطقی به نظر نمی رسد.

در این پژوهش سعی بر آن است که بر اساس پارامترهای تعیین شده برای بخش بندی مشتریان، تغییرات بازار در گذر زمان به تصویر کشیده شود و رفتار هر بخش و تغییرات آن مورد بررسی قرار بگیرد، به نحوی که هم ماهیت پویای رفتار افراد و جامعه مورد توجه قرار گیرد و هم بخش های بازار بر اساس ویژگی های مشتریان و محصولات تعیین شود.

استفاده از رویکرد داده کاوی پویا برای بخش بندی مشتریان به خوشه های رفتاری به جهت تحلیل رفتار مشتریان که در این تحقیق ارائه شده، بدیع و جدید بوده و تاکنون تحقیق مشابهی در این زمینه صورت نگرفته است.

۳.۱ اهمیت و ضرورت تحقیق

محیط کسب و کار امروزی بسیار رقابتی و پویاست و لازمه موفقیت در چنین محیطی انطباق با تغییرات است. بررسی‌ها نشان می‌دهد که تا کنون پژوهشگران از جنبه‌های مختلفی به موضوع بخش بندی بازار پرداخته‌اند و از متغیرها و تکنیک‌های متنوع و متفاوتی جهت پژوهش استفاده نموده‌اند. اما توجه به این نکته حائز اهمیت می‌باشد که نتایج این پژوهش‌ها زمانی کاربردی و مفید خواهند که ماهیت پویای جامعه در آن‌ها مورد توجه قرار گیرد. در میان روش‌ها و تکنیک‌هایی که تا کنون در این زمینه مورد استفاده قرار گرفته‌اند، داده‌کاوی به عنوان رایج‌ترین ابزار شناخته شده است. اما محدودیت مهمی که در این روش دیده می‌شود، فرض ثبات داده‌ها می‌باشد که سبب سهولت در کار داده‌کاوی می‌گردد ولی در بسیاری از کاربردهای دنیای واقعی صدق نمی‌کند. در این تحقیق این مهم مورد توجه قرار گرفته و با استفاده از رویکرد پویا در داده‌کاوی بخش بندی بصورت کارآمدتر انجام می‌شود.

۴.۱ هدف تحقیق

دستیابی به بخش بندی بازار^۸ از جمله بزرگترین دغدغه‌های شرکت‌های تولیدی و بازرگانی است. برای بخش بندی بازار ویژگی‌های متعددی می‌تواند مد نظر قرار گیرد که اولین و مهمترین متغیر، مشتری می‌باشد (میرواحدی و نوراله اوغلی، ۱۳۹۰). این تحقیق با ارائه روش خوشه بندی پویا برای بخش بندی مشتریان و تحلیل رفتار آن‌ها درصدد یافتن یک بخش بندی کارآمد و قابل اطمینان بر اساس تغییرات رفتار و ذائقه مشتریان شرکت می‌باشد.

⁸.Market Segmentation

۵.۱ سوالات تحقیق

سوال هایی که این تحقیق به دنبال پاسخ گویی به آن هاست عبارتند از:

۱. چگونه می توان با استفاده از تکنیک خوشه بندی فازی با رویکرد داده کاوی پویا تغییر رفتار مشتریان در فرایند زمانی را کشف نمود؟
۲. در فرایند زمانی ذائقه افراد (طبقات تخصیص داده شده) چگونه تغییر می نمایند؟
۳. آیا بین روش بخش بندی با رویکرد پویا و بخش بندی با رویکرد ایستا تفاوت وجود دارد؟ کدام روش کارآمدتر می باشد؟

۶.۱ روش تحقیق

تحقیق حاضر با توجه به ماهیت و هدف آن از نوع تحقیقات کاربردی و از جهت نحوه گردآوری داده ها از نوع توصیفی- پیمایشی می باشد.

۷.۱ قلمرو تحقیق

۱.۷.۱ قلمرو موضوعی

تحقیق حاضر از نظر چارچوب موضوعی جز تحقیقات بازاریابی محسوب می شود و نتایج حاصل از آن در تدوین استراتژی های مربوط به طراحی و توسعه محصول مورد استفاده قرار خواهد گرفت و لذا از این جهت جز تحقیقات استراتژیک و راهبردی برای شرکت محسوب می شود.

۲.۷.۱ قلمرو مکانی

با توجه به این که داده های مورد نیاز جهت انجام پایان نامه از پایگاه داده های نمایندگی های سامسونگ در شهر تهران جمع آوری شده اند، قلمرو مکانی این تحقیق شهر تهران می باشد.

۳.۷.۱ قلمرو زمانی

زمان انجام تحقیق از تیر ماه ۱۳۸۹ تا تیرماه ۱۳۹۳ می باشد.

۸.۱ تعریف واژه ها و اصطلاحات

۱.۸.۱ داده کاوی

داده کاوی استخراج اطلاعات و دانش و کشف الگوهای پنهان از پایگاه داده های بسیار بزرگ و پیچیده می باشد.

۲.۸.۱ خوشه بندی

خوشه بندی به عنوان یکی از فعالیت های داده کاوی می باشد و به گروه بندی کردن تراکنش ها و مشاهدات یا حالت ها در کلاس های مشابه می پردازد. همچنین یک خوشه مجموعه ای از رکوردها است که به هم شبیه می باشند و از رکوردهای بیرون خوشه تفاوت دارند.

۳.۸.۱ داده کاوی پویا

داده کاوی پویا به معنای درک و پیش بینی چگونگی تغییر الگوها در هر زمان است. برای نظارت بر تغییرات، مفاهیم داده کاوی پویا توسعه داده شده اند.

۴.۸.۱ منطق فازی

نظریه فازی برای بیان و تشریح عدم قطعیت و عدم دقت در رویدادها به وجود آمده است که براساس منطق چند ارزشی بوجود آمده است. پروفیسور لطفی زاده برای اولین بار تئوری مجموعه های فازی و منطق فازی^۹ را مطرح کرد، وی با معرفی نظریه مجموعه های فازی مقدمات مدل سازی اطلاعات نادقیق و استدلال تقریبی با معادله های ریاضی را فراهم نمود که در نوع خود تحولی عظیم در ریاضیات و منطق کلاسیک بوجود آورد.

^۹.Fuzzy logic

فصل دوم

ادبیات و پیشینه تحقیق

۱.۲ مقدمه

تحولات پیچیده و پرشتاب جهانی در عرصه‌های علم، فناوری، ارزش‌ها و معیارها بسیاری از سازمان‌های موفق جهان را بر آن داشته است تا اهداف، روش‌ها و ساختار خود را در جهت شناخت هر چه بیشتر مشتریان هدایت کنند. در چنین فضایی داشتن درک صحیح از مشتریان و فرایند مصرف، مزیت‌های متعددی را در بر دارد. این مزیت‌ها شامل کمک به مدیران در جهت تصمیم‌گیری، تهیه یک مبنای شناختی از طریق تحلیل رفتار مصرف‌کنندگان، کمک به قانون‌گذاران و تنظیم‌کنندگان برای وضع قوانین مربوط به خرید و فروش کالا و خدمات و در نهایت به مشتریان در جهت تصمیم‌گیری بهتر است. با علم به نحوه رفتار مخاطبان است که می‌توان رسانه و پیام مناسب را انتخاب کرد. بر این اساس، تحلیل رفتار مصرف‌کننده در مواردی همچون طراحی آمیخته بازاریابی، بخش بندی بازار و تعیین موقعیت و متمایز سازی محصول نیز امری لازم و حیاتی است. اما رفتار مشتری موضوعی بحث‌انگیز و چالشی است زیرا رفتار مشتری ماهیتی پویا دارد و در طول زمان تغییر می‌کند و این موضوع بر کیفیت تحلیل رفتار افراد تاثیر بسزایی دارد. لذا مدیران و بازاریابان باید در تصمیم‌گیری‌ها و برنامه‌ریزی‌های خود به این موضوع توجه داشته باشند و بر اساس آن زمینه را برای موفقیت هرچه بیشتر خود در فضای رقابتی را فراهم نمایند.

۲.۲ داده‌کاوی

داده‌کاوی به معنای یافتن نیمه خودکار الگوهای پنهان موجود در مجموعه داده‌های موجود می‌باشد. داده‌کاوی از مدل‌های تحلیلی، کلاس بندی و تخمین و برآورد اطلاعات و ارائه نتایج با استفاده از ابزارهای مربوطه بهره می‌گیرد. می‌توان گفت که داده‌کاوی در جهت کشف اطلاعات پنهان و روابط موجود در بین داده‌های فعلی و پیش‌بینی موارد نامعلوم و یا مشاهده نشده عمل می‌کند. برای انجام عملیات داده‌کاوی لازم است قبلاً روی داده‌های موجود پیش‌پردازش‌هایی انجام گیرد. عمل پیش‌پردازش

اطلاعات خود از دو بخش کاهش اطلاعات و خلاصه‌سازی و کلی‌سازی داده‌ها تشکیل شده است. کاهش اطلاعات عبارت است از تولید یک مجموعه کوچک تر، از داده‌های اولیه، که تحت عملیات داده‌کاوی نتایج تقریباً یکسانی با نتایج داده‌کاوی روی اطلاعات اولیه به دست دهد. پس از انجام عمل کاهش اطلاعات و حذف خصایص غیر مرتبط نوبت به خلاصه‌سازی و کلی‌سازی داده‌ها می‌رسد. داده‌های موجود در بانک‌های اطلاعاتی معمولاً حاوی اطلاعات در سطوح پایینی هستند، بنابراین خلاصه‌سازی مجموعه بزرگی از داده‌ها و ارائه آن به صورت یک مفهوم کلی اهمیت بسیار زیادی دارد. کلی‌سازی اطلاعات، فرآیندی است که تعداد زیادی از رکوردهای یک بانک اطلاعاتی را به صورت مفهومی در سطح بالاتر ارائه می‌نماید (هان و کمبر^{۱۰}، ۲۰۰۱).

در داده‌کاوی، چهار عمل اصلی انجام می‌شود که عبارتند از:

مدلسازی پیشگویی کننده، تقطیع پایگاه داده‌ها، تحلیل پیوند، تشخیص انحراف.

از عملیات‌های اصلی مذکور، یک یا بیش از یکی از آنها در پیاده‌سازی کاربردهای گوناگون داده‌کاوی استفاده می‌شوند.

نگاهی به ترجمه تحت‌اللفظی داده‌کاوی، به ما به درک بهت این واژه کمک می‌کند. Mine به معنای استخراج از منابع نهفته و با ارزش زمین اطلاق می‌شود. پیوند این کلمه با کلمه داده، جستجوی عمیق جهت پیدا کردن اطلاعات اضافی مفید که قبلاً نهفته بودند، از داده‌های قابل دسترس حجیم، را پیشنهاد می‌کند. داده‌کاوی یک رشته نسبتاً جدید علمی می‌باشد که از انجام تحقیقات در رشته‌های آمار، یادگیری ماشینی، علوم کامپیوتر خصوصاً مدیریت پایگاه داده شکل گرفته است (برسون و همکاران^{۱۱}، ۲۰۰۴).

¹⁰ . Han and Kamber

¹¹ . Berson, Smith and Thearling

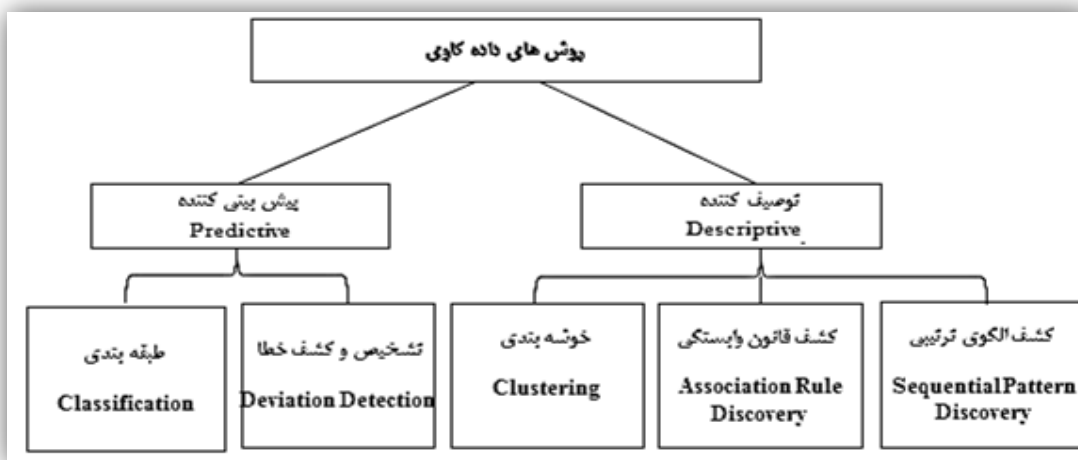
تعاریف متنوعی از داده‌کاوی در مراجع مختلف و توسط افراد مختلف ارائه شده است از جمله:

۱. داده‌کاوی فرایند شناخت الگوهای معتبر، جدید، ذاتا مفید و قابل فهم از داده‌ها می باشد.
 ۲. داده‌کاوی (کشف دانش از پایگاه داده‌ها نامیده می شود) نشانگر فرایند جالب استخراج دانش از قبل ناشناخته (الگو) از داده است.
 ۳. اصطلاح داده‌کاوی به فرایند نیم خودکار تجزیه و تحلیل پایگاه داده‌های بزرگ به منظور یافتن الگوهای مفید اطلاق می شود (ساندر^{۱۲}، ۲۰۰۳).
 ۴. فرایند کشف الگوهای مفید از داده‌ها را داده‌کاوی می گویند.
 ۵. فرایند انتخاب، کاوش و مدل بندی داده‌های حجیم، جهت کشف روابط نهفته با هدف به دست آوردن نتایج واضح و مفید، برای مالک پایگاه داده‌ها را، داده‌کاوی گویند.
- اما تعریفی که در اکثر مراجع به صورت مشترک ذکر شده عبارت است از " استخراج اطلاعات و دانش و کشف الگوهای پنهان از پایگاه داده‌های بسیار بزرگ و پیچیده".
- داده‌کاوی فرایندی است که از ابزارهای تحلیلی گوناگون برای کشف الگوها و روابط بین داده‌ها استفاده می کند که ممکن است برای اعتبار بخشیدن به پیش بینی استفاده شود (بری و لینوف^{۱۳}، ۱۹۹۷).
- داده‌کاوی یک متدولوژی بسیار قوی و با پتانسیل بالا می باشد که به سازمان کمک می کند که بر روی مهمترین اطلاعات از مخزن داده‌های خود تمرکز نمایند.
- داده‌کاوی کمک می کند تا سازمان‌ها با کاوش بر روی داده‌های یک سیستم، الگوها، روندها و رفتارهای آینده را کشف و پیش بینی کرده و بهتر تصمیم بگیرند. داده‌کاوی با استفاده از تحلیل وقایع گذشته یک تحلیل اتوماتیک و پیش بینانه ارائه می نماید و به سوالاتی جواب می دهد که پاسخ آن‌ها در گذشته ممکن نبوده و یا به زمان زیادی نیاز داشت.

¹² . Sander

¹³ . Berry and Linoff

ابزارهای داده‌کاوی الگوهای پنهانی را کشف و پیش‌بینی می‌کنند که متخصصان ممکن است به دلیل اینکه این اطلاعات و الگوها خارج از انتظار آن‌ها باشد، آن‌ها را مدنظر قرار ندهند و به آن‌ها دست نیابند.^{۱۴}



شکل (۱.۲): روش‌های داده‌کاوی^{۱۵}

۳.۲ خوشه‌بندی

خوشه‌بندی به عنوان یکی از فعالیت‌های داده‌کاوی می‌باشد و به گروه‌بندی کردن تراکنش‌ها و مشاهدات یا حالت‌ها در کلاس‌های مشابه می‌پردازد. همچنین یک خوشه مجموعه‌ای از رکوردها است که به هم شبیه می‌باشند و از رکوردهای بیرون خوشه تفاوت دارند. در خوشه‌بندی متغیر هدف وجود ندارد و به طبقه‌بندی تخمین و پیشگویی مقدار متغیر هدف نمی‌پردازد (لاروس^{۱۶}، ۲۰۰۵).

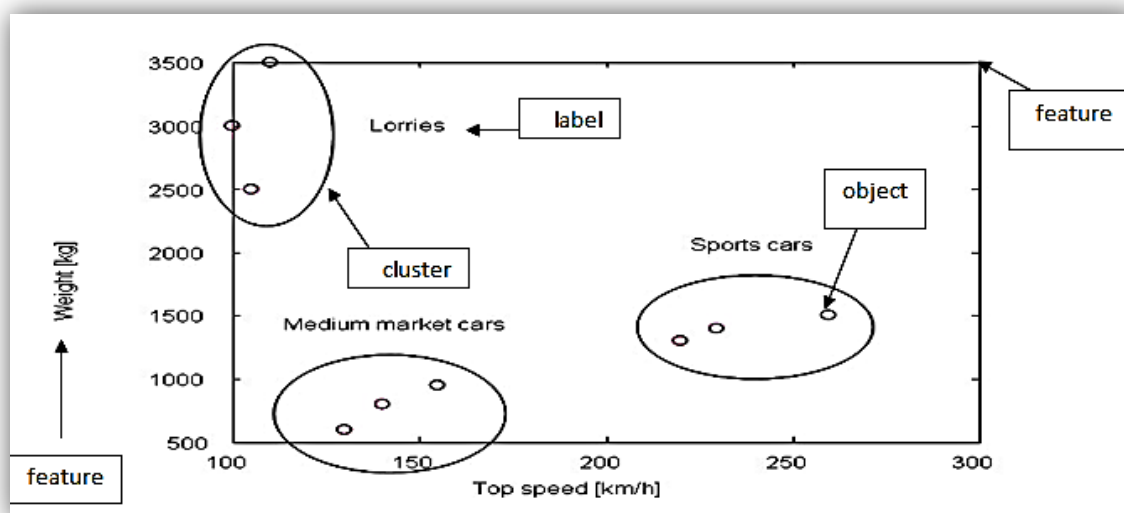
خوشه‌بندی یکی از شاخه‌های یادگیری بدون نظارت می‌باشد و فرآیند خودکاری است که در طی آن، نمونه‌ها به دسته‌هایی که اعضای آن مشابه یکدیگر می‌باشند تقسیم می‌شوند که به این دسته‌ها خوشه گفته می‌شود. بنابراین خوشه مجموعه‌ای از اشیاء می‌باشد که در آن اشیاء با یکدیگر مشابه بوده و با اشیاء

¹⁴ . <http://www.thearling.com/> an introduction to data mining.htm

¹⁵ . Comex2010-FUMworkshops-DataMining-13881029_54500.ppt

¹⁶ . Larose

موجود در خوشه‌های دیگر غیر مشابه می باشند. برای مشابه بودن می توان معیارهای مختلفی را در نظر گرفت مثلا می توان معیار فاصله را برای خوشه‌بندی مورد استفاده قرار داد و اشیائی را که به یکدیگر نزدیک تر هستند را بعنوان یک خوشه در نظر گرفت که به این نوع خوشه‌بندی، خوشه‌بندی مبتنی بر فاصله نیز گفته می شود. بعنوان یک مثال شکل زیر را در نظر بگیرید در این شکل هر یک از دایره‌های کوچک یک وسیله نقلیه (شیء) را نشان می دهد که با ویژگی‌های وزن و حداکثر سرعت مشخص شده اند. هر یک از بیضی‌ها یک خوشه می باشد و عبارت کنار هر بیضی برچسب آن خوشه را نشان می دهد. کل دستگاه مختصات که نمونه‌ها در آن نشان داده شده اند را فضای ویژگی می گویند.



شکل (۲-۲) : خوشه بندی وسایل نقلیه

همان طور که در شکل می بینید وسایل نقلیه به سه خوشه تقسیم شده اند. برای هر یک از این خوشه‌ها می توان یک نماینده در نظر گرفت مثلا می توان میانگین وسایل نقلیه باری را محاسبه کرد و به عنوان نماینده خوشه وسایل نقلیه باری معرفی نمود. در واقع الگوریتم های خوشه بندی اغلب بدین گونه اند که یک سری نماینده اولیه برای نمونه های ورودی در نظر گرفته می شود و سپس از روی میزان تشابه نمونه ها با این نماینده‌های مشخص می شود که نمونه به کدام خوشه تعلق دارد و بعد از این مرحله نماینده‌های

جدید برای هر خوشه محاسبه می شود و دوباره نمونه‌ها با این نماینده‌ها مقایسه می شوند تا مشخص شود که به کدام خوشه تعلق دارند و این کار آنقدر تکرار می شود تا زمانی که نماینده‌های خوشه‌ها تغییری نکنند.

۱.۳.۲ هدف از خوشه‌بندی

هدف خوشه‌بندی یافتن خوشه‌های مشابه از اشیاء در بین نمونه‌های ورودی می باشد اما چگونه می توان گفت که یک خوشه‌بندی مناسب است و دیگری مناسب نیست؟ می توان نشان داد که هیچ معیار مطلقی برای بهترین خوشه بندی وجود ندارد بلکه این بستگی به مساله و نظر کاربر دارد که باید تصمیم بگیرد که آیا نمونه‌ها بدرستی خوشه‌بندی شده‌اند یا خیر. با این حال معیارهای مختلفی برای خوب بودن یک خوشه بندی ارائه شده است که می تواند کاربر را برای رسیدن به یک خوشه‌بندی مناسب راهنمایی کند که در بخش‌های بعدی چند نمونه از این معیارها آورده شده است. یکی از مسایل مهم در خوشه بندی انتخاب تعداد خوشه‌ها می باشد. در بعضی از الگوریتم‌ها تعداد خوشه‌ها از قبل مشخص شده است و در بعضی دیگر خود الگوریتم تصمیم می گیرد که داده‌ها به چند خوشه تقسیم شوند (قاسمی و خانگلدی، ۱۳۸۸).

۲.۳.۲ انواع خوشه‌بندی

روش‌های خوشه‌بندی بسیار متنوع بوده و تفکیک آن‌ها از یکدیگر بسیار مهم است. انتخاب بهترین این روش‌ها برای رسیدن به بهترین نتیجه از بحث‌های مهم در این زمینه است. متأسفانه روش مناسبی که بتواند بهترین روش را از بین روش‌های متنوع خوشه‌بندی انتخاب کند وجود ندارد. در اکثر موارد نیز تنها آشنایی محقق و قابل دسترس بودن نرم افزارهای مربوطه است که مبنای استفاده از روش می شود.

روش‌های خوشه‌بندی متفاوتی وجود دارد که برای تمایز آن‌ها از یکدیگر از پنج خصوصیت استفاده می‌شود. ابتدا به تعریف این خصوصیات پرداخته سپس روش‌های خوشه‌بندی را بر اساس آن‌ها معرفی می‌کنیم.

- **روش‌های انحصاری در مقابل روش‌های غیر انحصاری**^{۱۷}: روش‌های انحصاری روش‌هایی هستند که هر گونه تنها به یک گروه تعلق می‌گیرد در حالی که روش‌های غیر انحصاری یک فرد به یک یا چند گروه تعلق می‌گیرد. روش‌های کلاسیک خوشه‌بندی از جمله روش‌های انحصاری و روش‌های خوشه‌بندی فازی از جمله روش‌های غیر انحصاری هستند.
- **روش‌های تکرار توالی در مقابل روش‌های همزمان**^{۱۸}: در روش‌های تکرار توالی عمل خوشه‌بندی در دفعات متفاوت بر روی افراد انجام می‌گیرد در حالی که در روش‌های همزمان این عمل به طور آنی بر روی همه افراد انجام می‌شود.
- **روش‌های سلسله مراتبی در مقابل روش‌های غیر سلسله مراتبی**^{۱۹}: در روش‌های سلسله‌مراتبی خوشه‌ها به شکل سلسله‌مراتبی از خوشه‌های بزرگ تا کوچک و یا برعکس تعیین می‌شوند. به عبارت دیگر در این روش‌ها افراد در یک خوشه قرار می‌گیرند که خود این خوشه نیز به خوشه‌های دیگری تقسیم می‌شود. در روش‌های غیر سلسله‌مراتبی افراد به طور مستقیم در خوشه‌های متفاوتی قرار می‌گیرند.
- **روش‌های تجمعی در مقابل روش‌های مقسمی**^{۲۰}: در روش‌های تجمعی شبیه‌ترین گونه‌ها در یک خوشه قرار گرفته و این خوشه‌ها با خوشه‌های دیگر تلفیق شده تا خوشه بزرگتری را بوجود آورند و این عمل تا رسیدن به بزرگترین خوشه که همه گونه‌ها در آن قرار دارند ادامه می‌یابد. در

¹⁷ . Exclusive versus Non-exclusive

¹⁸ . Recrusive sequence versus Simultaneous

¹⁹ . Heirachal versus Non-heirachical

²⁰ . Agglomerative versus divisive

روش‌های مقسمی ابتدا گونه‌ها به دو خوشه تقسیم شده و سپس این خوشه‌ها به خوشه‌های ریزتری تقسیم می‌شوند.

• **روش‌های چند صفتی در مقابل روش‌های تک صفتی**^{۲۱}: برای قراردادن گونه‌ها در خوشه‌های متفاوت در روش‌های چند صفتی از تمامی خصوصیات (متغیرها) آن‌ها استفاده می‌شود در صورتی که در روش‌های تک صفتی تنها از یک ویژگی (متغیر) برای خوشه‌بندی استفاده می‌شود (طهماسبی، ۱۳۹۰).

فرآیند خوشه‌بندی سعی دارد که یک مجموعه داده را به چندین خوشه تقسیم نماید بطوری که داده‌های قرار گرفته در یک خوشه با یکدیگر شبیه بوده و با داده‌های خوشه‌های دیگر متفاوت باشند. در حال حاضر روش‌های متعددی برای خوشه‌بندی داده‌ها وجود دارد که بر اساس نوع داده‌ها، شکل خوشه‌ها، فاصله داده‌ها و غیره عمل خوشه‌بندی را انجام می‌دهند. مهمترین روشهای خوشه‌بندی در زیر معرفی شده‌اند:

• **روش‌های تقسیم‌بندی (تفکیکی)**: روش‌های خوشه‌بندی که به روش تقسیم بندی عمل می‌کنند، داده‌های موجود در یک مجموعه داده را به k خوشه تقسیم می‌کنند، بطوری که هر خوشه دو خصوصیت زیر را داراست:

- هر خوشه یا گروه حداقل شامل یک داده می‌باشد.
- هر داده موجود در مجموعه داده دقیقاً به یک گروه یا خوشه تعلق دارد.
- معیار اصلی در چنین مجموعه داده‌هایی میزان شباهت داده‌های قرار گرفته در هر خوشه می‌باشد. در حالی که داده‌های قرار گرفته در دو خوشه مختلف از نظر شباهت با یکدیگر فاصله

²¹. Pplythetic versus Monothetic

زیادی دارند. مقدار k که بعنوان پارامتر استفاده می‌گردد، هم می‌تواند بصورت پویا تعیین گردد و هم اینکه قبل از شروع الگوریتم خوشه‌بندی مقدار آن مشخص گردد.

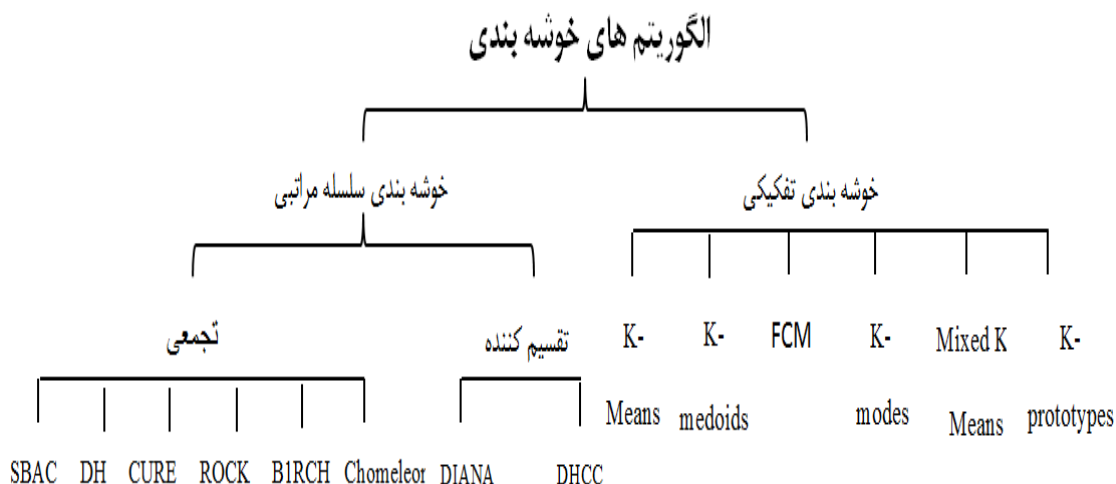
- **روش‌های سلسله مراتبی :** روش‌های سلسله مراتبی به دو دسته کلی روش‌های bottom-up و روش‌های top-down تقسیم می‌گردند. روش‌های سلسله مراتبی bottom-up به این صورت عمل می‌کنند که در شروع هر کدام از داده‌ها را در یک خوشه جداگانه قرار می‌دهد و در طول اجرا سعی می‌کند تا خوشه‌هایی نزدیک به یکدیگر را با هم ادغام نماید. این عمل ادغام تا زمانی که یا تنها یک خوشه داشته باشیم و یا اینکه شرط خاتمه برقرار گردد، ادامه می‌یابد. روش‌های top-down دقیقا به طریقه عکس عمل می‌کنند، به این طریق که ابتدا تمام داده‌ها را در یک خوشه قرار می‌دهد و در هر تکرار از الگوریتم، هر خوشه به خوشه‌های کوچکتر شکسته می‌شود و این کار تا زمانی ادامه می‌یابد که یا هر کدام از خوشه‌ها تنها شامل یک داده باشند و یا شرط خاتمه الگوریتم برقرار گردد. شرط خاتمه معمولا تعداد کلاستر یا خوشه می‌باشد.

- **روش‌های مبتنی بر چگالی :** اکثر روش‌های خوشه‌بندی که به روش تقسیم‌بندی عمل می‌کنند معمولا از تابع فاصله بعنوان تابع معیار خود بهره می‌برند. استفاده از چنین معیاری باعث می‌گردد که الگوریتم خوشه‌بندی تنها قادر به ایجاد خوشه‌هایی با اشکال منظم باشد. در صورتی که اگر خوشه‌های واقعی در داده‌ها دارای اشکال غیرمنظمی باشند، این الگوریتم‌ها در خوشه‌بندی آنها با مشکل مواجه می‌گردند. برای حل این گونه مشکلات یکسری از روشها برای خوشه‌بندی پیشنهاد گردیده‌اند که عمل خوشه‌بندی را بر مبنای چگالی داده‌ها انجام می‌دهند. ایده اصلی در این روشها بر این اساس است که خوشه‌ها تا زمانی که داده‌های قرار گرفته همسایگی خوشه‌ها از حد معینی بیشتر باشد، رشد می‌کنند و بزرگ می‌شوند. چنین روش‌هایی قادرند خوشه‌هایی با شکل‌های نامنظم نیز ایجاد نمایند.

البته دسته دیگری از روش‌های خوشه‌بندی مانند روش‌های مبتنی بر گرید، روش‌های مبتنی بر مدل و ... وجود دارند که می‌توانید آنها را در همین مرجع مطالعه نمایید (هان و کمبر، ۲۰۰۱).

۳.۳.۲ الگوریتم‌های خوشه‌بندی

تاکنون الگوریتم‌های مختلف و متعددی برای خوشه‌بندی ارائه شده‌اند که در ادامه به برخی از مهم‌ترین و پرکاربردترین آن‌ها اشاره می‌کنیم.



شکل (۲-۳): انواع الگوریتم‌های خوشه‌بندی

در اینجا ما قصد نداریم وارد مباحث مربوط به الگوریتم‌های خوشه‌بندی شویم و تنها روش K-means و FCM را به دلیل استفاده در پژوهش تشریح خواهیم نمود. در هر فرآیند خوشه‌بندی مراحل طی می‌شود که عبارتند از:

- تهیه و ارائه ماتریس داده‌ها
- استاندارد کردن ماتریس داده‌ها
- محاسبه ماتریس مجاورت (فاصله یا مشابهت)
- اجرای روش خوشه‌بندی

• محاسبه معیار اعتبار (حسین زاده و سلاجقه، ۲۰۱۲).

۴.۳.۲ معیار فاصله (تشابه)

یکی از مسائل مهم در خوشه بندی تعیین معیاری برای محاسبه فاصله بین داده هاست. معیارهای مختلفی برای اندازه گیری فاصله بین اشیا وجود دارد که از معمول ترین و پرکاربردترین آنها می توان فاصله اقلیدسی را نام برد. فاصله اقلیدسی برای دو نقطه X, Y در فضای n بعدی از رابطه زیر بدست می آید.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1-2)$$

اگر معیار تشابه در تابع هدف بر اساس فاصله تعریف شود می توان از تعاریف مختلفی که در مورد فاصله وجود دارد استفاده کرد که در زیر چند نمونه از این توابع آورده شده است:

Distance Function	Formula and Comments
Euclidean distance	$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Hamming (city block) distance	$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i - y_i $
Tchebyshev distance	$d(\mathbf{x}, \mathbf{y}) = \max_{i=1,2,\dots,n} x_i - y_i $
Minkowski distance	$d(\mathbf{x}, \mathbf{y}) = \sqrt[p]{\sum_{i=1}^n (x_i - y_i)^p}, p > 0$
Canberra distance	$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \frac{ x_i - y_i }{x_i + y_i}, x_i \text{ and } y_i \text{ are positive}$
Angular separation	$d(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n x_i y_i}{\left[\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 \right]^{1/2}}$

شکل (۴-۲): معیارهای تشابه بر اساس توابع فاصله مختلف

۵.۳.۲ الگوریتم C-means

این الگوریتم یکی از معروف‌ترین و ساده‌ترین الگوریتم‌ها است و علی‌رغم اینکه سال‌ها از ابداع آن می‌گذرد و پس از آن تعداد زیادی الگوریتم خوشه‌بندی توسعه داده شده‌اند، اما به دلیل مزایایی مثل سهولت پیاده‌سازی، سادگی و کارایی بالا هنوز هم به طور وسیعی مورد استفاده قرار می‌گیرد .

این روش علی‌رغم سادگی آن یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر (مانند خوشه‌بندی فازی) محسوب می‌شود. این روش روشی انحصاری و مسطح محسوب می‌شود (الپادین^{۲۲}، ۲۰۰۴).

برای الگوریتم C-means تابع زیر به عنوان تابع هدف مطرح است.

$$J = \sum_{j=1}^k \sum_{i=1}^x \|x_i^j - c_j\|^2 \quad (2-2)$$

که $\| \|$ معیار فاصله بین نقاط و c_j مرکز خوشه j ام است.

در این الگوریتم میانگین داده‌های یک خوشه، نماینده یک خوشه است. مراحل الگوریتم به صورت زیر است:

۱. انتخاب تعداد خوشه‌ها، C .
۲. انتخاب تصادفی C نقطه برای مقدار دهی اولیه مراکز خوشه‌ها.
۳. برای هر نمونه نزدیکترین مرکز خوشه را با استفاده از فاصله اقلیدسی پیدا می‌کنیم. در پایان این مرحله تمام نمونه‌ها در C خوشه قرار دارند.
۴. برای هر کدام از خوشه‌ها، مراکز ثقل جدید را محاسبه و مقدار آن را به روز رسانی می‌کنیم.
۵. تکرار مراحل ۲ و ۳ تا زمانی که الگوریتم خاتمه یابد.

²². Alpaydin

الگوریتم تا زمانی ادامه میابد که معیار مربع خطای تعریف شده به صورت رابطه زیر حداقل شود. حاصل این عبارت مجموع فاصله اشیا از خوشه خودشان است.

$$E = \sum_{i=1}^c \sum_{p \in c_i} (p - m_i)^2 \quad (3-2)$$

که در آن m_i مرکز خوشه i ام و C تعداد خوشه هاست.

همان طور که مشاهده شد هدف این الگوریتم حداقل کردن فاصله بین اجزای یک خوشه و حداکثر کردن فاصله بین اجزای خوشه های مجزاست. اما این الگوریتم معایبی دارد که از جمله آن می توان موارد زیر را برشمرد.

- حساس بودن نسبت به داده های دور از مرکز
- وابسته بودن نتایج به انتخاب مراکز اولیه و انجام یک جستجوی محلی
- ناتوانی در شناسایی داده نویز (حسین زاده و سلاجقه، ۲۰۱۲).

۶.۳.۲ الگوریتم FCM

مشابه الگوریتم C میانگین کلاسیک در این الگوریتم نیز تعداد خوشه ها (c) از قبل مشخص شده است. تابع هدفی که برای این الگوریتم تعریف شده است بصورت زیر می باشد:

$$J = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d_{ik}^2 = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|x_k - v_i\|^2 \quad (4-2)$$

در فرمول فوق m یک عدد حقیقی بزرگتر از ۱ است که در اکثر موارد برای m عدد ۲ انتخاب می شود. x_k نمونه k ام است و v_i نماینده یا مرکز خوشه i ام است. U_{ik} میزان تعلق نمونه i ام در خوشه k ام را نشان می دهد. علامت $\|*\|^2$ میزان تشابه (فاصله) نمونه با (از) مرکز خوشه می باشد که می توان از هر تابعی که بیانگر تشابه نمونه و مرکز خوشه باشد را استفاده کرد. از روی U_{ik} می توان یک ماتریس U تعریف کرد که دارای C سطر و n ستون می باشد و مولفه های آن هر مقداری بین ۰ تا ۱ را می توانند

اختیار کنند. اگر تمامی مولفه‌های ماتریس U بصورت ۰ و ۱ باشند الگوریتم مشابه C میانگین کلاسیک خواهد بود. با اینکه مولفه‌های ماتریس U می‌توانند هر مقداری بین ۰ تا ۱ را اختیار کنند اما مجموع مولفه‌های هر یک از ستون‌ها باید برابر ۱ باشد و داریم:

$$\sum_{i=1}^c u_{ik} = 1, \quad \forall k = 1, \dots, n \quad (5-2)$$

معنای این شرط این است که مجموع تعلق هر نمونه به C خوشه باید برابر ۱ باشد. با استفاده از شرط فوق و مینیمم کردن تابع هدف خواهیم داشت:

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)}} \quad (6-2)$$

9

$$v_i = \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m} \quad (7-2)$$

مراحل الگوریتم:

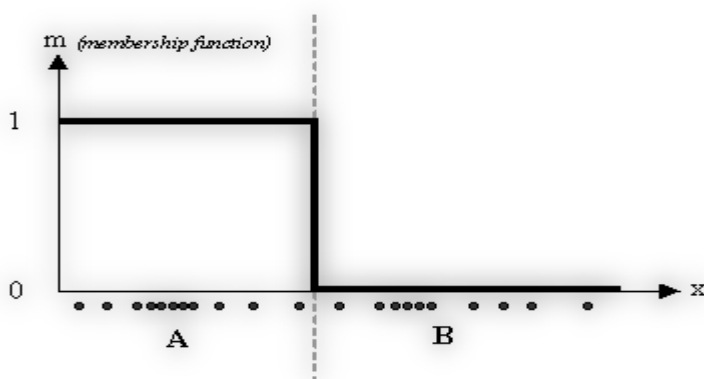
۱. مقدار دهی اولیه برای c , m و U^0 . خوشه‌های اولیه حدس زده شوند.
۲. مراکز خوشه‌ها محاسبه شوند (محاسبه v_i ها).
۳. محاسبه ماتریس تعلق از روی خوشه‌های محاسبه شده در ۲.
۴. اگر $\|U^{l+1} - U^l\| \leq \varepsilon$ الگوریتم خاتمه می‌یابد و در غیر اینصورت برو به مرحله ۲.

برای مشاهده عملکرد خوشه بندی فازی به مثال زیر توجه کنید. در شکل زیر یک توزیع یک بعدی از نمونه‌های ورودی را آورده شده است.



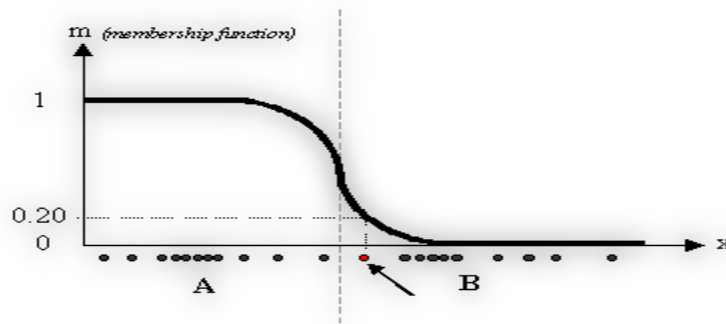
شکل (۲-۵): توزیع یک بعدی نمونه‌ها

اگر از الگوریتم C میانگین کلاسیک استفاده کنیم داده‌های فوق به دو خوشه مجزا تقسیم خواهند شد و هر نمونه تنها متعلق به یکی از خوشه‌ها خواهد بود. بعبارت دیگر تابع تعلق هر نمونه مقدار ۰ یا ۱ خواهد داشت. نتیجه خوشه بندی کلاسیک مطابق شکل زیر است:



شکل (۲-۶): خوشه بندی کلاسیک نمونه‌های ورودی

شکل ۶ تابع تعلق مرتبط به خوشه A را نشان می‌دهد. تابع تعلق خوشه B متمم تابع تعلق A می‌باشد. همانطور که مشاهده می‌کنید نمونه‌های ورودی تنها به یکی از خوشه‌ها تعلق دارند و بعبارت دیگر ماتریس U بصورت باینری می‌باشد. حال اگر از خوشه بندی فازی استفاده کنیم خواهیم داشت:



شکل (۷-۲): خوشه بندی فازی نمونه ها

مشاهده می کنید که در این حالت منحنی تابع تعلق هموارتر است و مرز بین خوشه ها بطور قطع و یقین مشخص نشده است. بعنوان مثال نمونه ای که با رنگ قرمز مشخص شده است با درجه تعلق ۰.۲ به خوشه A و با درجه تعلق ۰.۸ به خوشه B نسبت داده شده است.

نقاط قوت الگوریتم C میانگین فازی:

- همیشه همگرا می شود.
- بدون نظارت بودن الگوریتم.

نقاط ضعف الگوریتم C میانگین فازی:

- زمان محاسبات زیاد است.
- حساس به حدس های اولیه می باشد و ممکن است در مینیمم های محلی متوقف شود.
- حساس به نویز می باشد (قاسمی و خانگلدی، ۱۳۸۸).

خوشه‌بندی فازی یک روش خوشه‌بندی مشترک است که اجازه می‌دهد عناصر به بیش از یک خوشه بطور همزمان تعلق داشته باشند بر خلاف خوشه بندی کلاسیک که نتیجه آن، خوشه های ناسازگار (مانعه الجمع) هستند. دلایلی که برای اتخاذ روش خوشه‌بندی فازی وجود دارد عبارتند از:

- به راحتی با توزیع روش‌های بهینه‌سازی اختیاری سازگار می‌شود.
- به دلیل دشواری شناسایی مرز شفاف بین خوشه‌ها در مسائل دنیای واقعی، خوشه‌های فازی که ظاهر می‌شوند، جذاب‌تر از خوشه‌های قطعی می‌باشند.
- خوشه‌بندی فازی کمتر دچار مشکل تعیین مناسب‌ترین محل می‌شود.
- الگوریتم خوشه‌بندی فازی در محاسبات کاراتر است زیرا در روش‌های تخمینی با احتمال کمتری تغییر در مقادیر عضویت خوشه‌ها رخ می‌دهد.
- در پایان خوشه‌بندی فازی مقدار عضویت برای هر مجموعه از پاسخ‌ها نشان می‌دهد که آیا خوشه دیگری وجود دارد که نتایجی به خوبی این خوشه‌ها داشته باشد؟
- روش خوشه بندی کلاسیک این نتیجه را کشف نمی‌کند (ان و دورسو^{۲۳}، ۲۰۱۱).

۴.۲ طبقه بندی

فرایند طبقه‌بندی در واقع نوعی یادگیری با ناظر می‌باشد که در طی دو مرحله انجام می‌گردد. در مرحله اول مجموعه‌ای از داده‌ها که در آن هر داده شامل تعدادی خصوصیت دارای مقدار و یک خصوصیت بنام خصوصیت کلاس می‌باشد، برای ایجاد یک مدل داده بکار می‌روند که این مدل داده در واقع توصیف کننده مفهوم و خصوصیات مجموعه داده‌هایی است که این مدل از روی آن‌ها ایجاد شده است. مرحله دوم فرایند طبقه‌بندی اعمال یا بکارگیری مدل داده ایجاد شده بر روی داده‌هایی است که شامل تمام

²³ . Ann and D'Urso

خصوصیات داده‌هایی که برای ایجاد مدل داده بکار گرفته شده‌اند، می‌باشد، بجز خصوصیت کلاس این مقادیر که هدف از عمل طبقه‌بندی نیز تخمین مقدار این خصوصیت می‌باشد.

الگوریتم‌ها و روش‌های مختلفی برای طبقه‌بندی تاکنون پیشنهاد شده‌اند که برای مثال می‌توان از روش‌های طبقه‌بندی با استفاده از درخت تصمیم، طبقه‌بندی بیزین، SVM، طبقه‌بندی با استفاده از شبکه‌های عصبی، طبقه‌بندی مبتنی بر قواعد و ... نام برد (میشل^{۲۴}، ۱۹۹۷).

۵.۲ کشف قواعد وابستگی

بحث قواعد وابستگی به مقوله کشف عناصری یا المان‌هایی در یک مجموعه داده می‌پردازد که معمولاً با یکدیگر اتفاق می‌افتند و به عبارتی رخداد آن‌ها به نوعی با یکدیگر ارتباط دارد. بطور کلی هر قاعده یا rule که از این مجموعه داده بدست می‌آید، دارای شکل کلی بصورت $X \rightarrow Y$ می‌باشد که نشان می‌دهد چنانچه الگوی X اتفاق بیفتد، با احتمال بالایی الگوی Y نیز اتفاق خواهد افتاد (هان و کمبر، ۲۰۰۱).

۶.۲ کشف الگوی ترتیبی

کشف الگوی ترتیبی تشابه زیادی به کشف قوانین انجمنی دارد و هر دو شامل یک مجموعه قلم داده هستند با این تفاوت که مدل‌های ترتیبی انتقال‌های بین قلم داده‌ها را تحلیل می‌کند، در حالی که در مدل‌های انجمنی فرض بر این است که تراکنش‌های مختلف یک مشتری همگی مستقل از همدیگرند. در الگوی ترتیبی خرید یک صفحه کلید قبل از خرید یک مشواره با خرید یک مشواره قبل از صفحه کلید متفاوت است. به طور کلی داده‌های ورودی به روش‌های کشف الگوی ترتیبی شامل لیستی از تراکنش‌ها به همراه زمان رخ داد آنهاست^{۲۵}.

²⁴ . Mitchell

²⁵ . <http://ba-it.mihanblog.com/>

۷.۲ داده‌کاوی پویا

داده‌کاوی پویا در سال‌های گذشته توجه بسیاری را به خود جلب کرده است. در داده‌کاوی پویا بیان می‌شود که بر خلاف داده‌کاوی ایستا، در بسیاری از کاربردهای دنیای واقعی مثل رفتار خرید مشتریان، تغییرات ساختار داده مشاهده می‌شود. به عبارت دیگر در داده‌کاوی ایستا مسئله کشف الگوهای اصلی در مجموعه داده معین است ولی در داده‌کاوی پویا مسئله درک و پیش‌بینی چگونگی تغییر الگوها در هر زمان است.

الگوریتم‌های خوشه‌بندی کلاسیک ساختار داده‌ها را ایستا فرض می‌کنند که در طول زمان تغییر نمی‌کند. هر چند مکرراً دیده شده که تغییرات ساختار داده در بیشتر زمان‌ها تحلیل می‌شوند. برای نظارت بر تغییرات، مفاهیم داده‌کاوی پویا توسعه داده شده‌اند به ویژه برای خوشه‌بندی پویا. هدف اصلی این روش‌ها این است که مدل‌های اصلی را هر زمان که تغییری در ترکیب داده‌ها مشاهده شود، به روز-رسانی کنند. برای این کار دو مرحله باید انجام شود:

اول، کشف تغییر در ترکیب داده‌ها که نیاز به تعریف معیاری برای شناسایی تغییرات دارد. دوم، به روزرسانی مدل‌های داده‌کاوی بر اساس آن. معمولاً این کار با به روزرسانی پارامترهای مدل‌های اولیه انجام می‌شود (پیتز و همکاران^{۲۶}، ۲۰۱۲).

در مقایسه با داده‌کاوی ایستا، داده‌کاوی پویا ظرفیت بیشتری برای سروکار داشتن با داده‌های نامطمئن، بی‌قاعده و وابسته به زمان دارد..

همان‌گونه که ذکر گردید در داده‌کاوی کلاسیک فرض بر این است که مجموعه داده‌هایی که قرار است هر یک از اعمال یا روش‌های داده‌کاوی بر روی آنها اعمال گردد، بصورت ایستا در داخل یک پایگاه یا در

²⁶ . Peters ,Weber , Nowatzke

یک فایل وجود دارند و در حین عمل داده‌کاوی نیز این داده‌ها تغییری نمی‌کنند. این گونه فرض‌ها انجام عمل داده‌کاوی را بسیار ساده‌تر می‌نماید. اما نکته‌ای که باید در اینجا مورد توجه قرار گیرد این مطلب است که موارد ذکر شده همچون عدم تغییر داده‌ها، در دسترس بودن تمام داده‌ها و ... در بسیاری از کاربردهای دنیای واقعی صدق نمی‌کند، بلکه داده‌های این کاربردها معمولاً همواره در حال تغییر و تحول هستند و از طرفی حجم این داده‌ها نیز به قدری زیاد است که امکان ذخیره‌سازی آنها نیز وجود ندارد. بنابراین تنها قسمت کوچکی از داده‌ها در هر لحظه در اختیار ما خواهد بود، لذا برای داده‌کاوی بر روی داده‌هایی از این دست، نمی‌توان روش‌های سنتی و قدیمی داده‌کاوی را بکار برد بلکه این نوع داده‌ها مجموعه داده‌ها نیاز به الگوریتم‌ها و روش‌های داده‌کاوی پویایی دارند که از پس ویژگی‌های این داده‌ها برآیند. بنابراین نیاز به نوعی داده‌کاوی پویا احساس می‌شود. بنابراین منظور ما از داده‌کاوی پویا داده‌کاوی بر روی داده‌هایی است که همواره در حال تغییر و تحول بوده و حجم این داده‌ها بقدری زیاد است که نمی‌توان آنها را در یک منبع ذخیره‌سازی ذخیره کرد و تنها بخش کوچکی از داده‌ها در هر لحظه در دسترس می‌باشد (هان و کمبر، ۲۰۰۱).

۸.۲ خوشه بندی پویا

یکی از اولین روش‌های خوشه بندی پویا در دیدی (۱۹۷۳)^{۲۷} ارائه شده است که در آن گروهی از اشیا که "نمونه"^{۲۸} نامیده می‌شوند، با خوشه‌هایی که از آن‌ها استخراج می‌شود، در طول زمان سازگار می‌شوند. روش‌های خوشه بندی پویا با توجه به ماهیت داده ورودی برای توصیف این گونه ترکیب‌ها مورد توجه قرار می‌گیرند، خواه داده‌های مربوطه ایستا باشند یا پویا. داده‌های ورودی در صورتی ایستا هستند که به تغییرات زمان وابسته نباشند و نقطه مقابل آن‌ها داده‌های ورودی پویا هستند. چندین سیستم خوشه

²⁷.Diday

²⁸.Sampelings

بندی ارائه شده است که با ورودی‌های ایستا سروکار دارند ولی در خلال خوشه بندی از عناصر پویا استفاده می‌کنند. به عبارت دیگر پویا سازی الگوریتم مربوطه در حین اعمال در مجموعه داده ایستا، انجام می‌شود. چملون^{۲۹} نمونه ای از این سیستم هاست که در خوشه بندی سلسله مراتبی بکار برده می‌شود. در این نوع خوشه بندی در هر مرتبه تصمیم‌گیری به طور پویا بر ویژگی‌های خوشه فعلی منطبق می‌شود (کریپس و همکاران ۱۹۹۹).

به جز آن چه که در خوشه بندی سلسله مراتبی ذکر شد، در روش‌های خاص خوشه بندی کاربردی از قبیل C-means و Fuzzy C-means قبل از اجرای الگوریتم باید تعداد خوشه‌ها تعیین گردند. تعیین تعداد مناسب خوشه‌ها موضوع چندین پژوهش نیز بوده است (بزدک و همکاران ۱۹۹۹). در "خوشه بندی تفکیکی پویا با استفاده از استراتژی‌های تکامل" تعداد خوشه‌ها در حین اجرا با استفاده از الگوریتم‌های تکاملی بهینه می‌شود (لی و انتنسون ۲۰۰۱).

در روش‌های خوشه بندی‌ای که در بالا به آن‌ها اشاره شد از عناصر پویا در حین اجرای الگوریتم روی مجموعه داده ایستا استفاده می‌شود. در حالی که ما وضعیت داده‌های پویا را در دو حالت متمایز مورد تحلیل قرار خواهیم داد. در یک حالت خوشه بندی می‌تواند با توجه به رشد مقادیر ویژگی انجام شود که در این صورت خوشه بندی پویا یعنی، خوشه بندی مسیرها نه خوشه بندی بردار ویژگی‌ها. در حال دوم قصد ما این است که دریابیم ترکیب کلاس‌ها در طول زمان ممکن است چه تغییری کند و در صورتی که اطلاعات جدید در دسترس باشد، یک کلاس چطور به روزرسانی خواهد شد. در ادامه هر دو مورد به طور دقیق مورد مطالعه قرار خواهند گرفت.

در شرایطی که مقدار ویژگی‌های فعلی برای توصیف موضوع موردنظر کافی نباشند، تحلیل مسیر ویژگی‌ها بسیار مناسب خواهد بود. مثلا در علوم پزشکی شرایط بیمار نه تنها به فشار خون بلکه به

²⁹.Chameleon

بهبودی وی در قیاس با شرایط گذشته نیز بستگی دارد. مثال دیگر در این زمینه پیش بینی قیمت سهام می باشد. زمانی که برای پیش بینی افزایش قیمت در آینده، اطلاعات ارزش فعلی سهام کافی نباشد. در این حالت باید به قیمت سهام های مرتبط گذشته برای پیش بینی دقت کنیم.

روشی که در آن مسیر ویژگی خوشه بندی شده است، روش ماتریوشکا^{۳۰} می باشد (لی و همکاران^{۳۱} ۲۰۰۲) که بر اساس مدل مخفی مارکو (HHM)^{۳۲} برای خوشه بندی داده های موقتی طراحی شده است. این مدل با استفاده از داده های ورودی موقتی، مقدار مطلوب کلاس ها را توضیح می دهد. هر کلاس به عنوان یک (HHM) مشخص می شود و اعضای مربوط به هر کلاس را تخصیص می دهد.

یک سیستم ارزیابی و استنتاج عصبی فازی برای برآورد سری های زمانی پویا (DENFIS)^{۳۳} استفاده است (کاسابو و سانگ^{۳۴} ۲۰۰۲). بر اساس روش خوشه بندی استنتاجی (ECM)^{۳۵} نوعی مجموعه قواعد فازی به طور پویا ایجاد می گردد.

منطق فازی نیز توانایی خود را در خوشه بندی جریان داده^{۳۶} اثبات کرده است. یک جریان داده در واقع یک توالی از داده هایی بشکل (x_1, \dots, x_n) می باشد، که هر کدام از x_i ها یک خصوصیت دارای مقدار می باشند. خوشه بندی فازی مزایای بسیاری دارد زیرا به جای آن که تغییرات خوشه ها زود به زود رخ دهد، بسیار آرام انجام می شود.

۹.۲ روش های نوین برای خوشه بندی فازی پویا

در جدول زیر یک دسته بندی از حالات های مختلف خوشه بندی درحالی که داده های ورودی پویا

³⁰ .Matryoshka

³¹ . Li, Biswas , Dale, and Dale

³² .Hidden Markov model (HHM)

³³ .Dynamic evolving neural fuzzy inference system

³⁴ . Kasabov and Song

³⁵ .Evolving clustering method

³⁶ .Data stream

باشند تهیه شده است. در کلاس‌های ایستا خوشه بندی فقط در یک نقطه از زمان انجام می شود و ترکیب کلاس‌های به روز رسانی نمی شود. اگر اعضای اضافه شده ایستا باشند، به عبارت دیگر؛ اگر بردار ویژگی که براساس مقادیر واقعی توصیف شده ، شامل مقادیر ویژگی فعلی استفاده شده باشد، آن گاه یک نمونه از خوشه بندی کلاسیک وجود خواهد داشت. اگر بخواهید رشد مقادیر قبلی ویژگی را در خوشه بندی بررسی کنید، اعضای پویا را به کلاس‌های ایستا اختصاص دهید. در این حالت بردار ویژگی شامل مسیر است به جای مقادیر واقعی.

در کلاس‌های پویا، ترکیب کلاس‌ها باید مکررا به روز رسانی شوند. اگر در حال به روز رسانی با اعضای ایستا باشید آن گاه بردارهای ویژگی مربوطه با مقادیر فعلی ایجاد می شود. اختصاص اعضای پویا به کلاس‌های پویا، نمونه ای از به روز رسانی کلاس‌ها با مسیر ویژگی است.

در قسمت زیر روش‌های توسعه یافته خوشه بندی فازی برای موارد (کلاس ایستا- اعضا پویا) و (کلاس پویا - اعضا ایستا) ارائه شده است. همچنین نشان داده شده است که کارکردن با منطق فازی تهیه شده مزایای ویژه‌ای در خوشه بندی پویا دارد، به این دلیل که درجه عضویت مربوطه یک ابزار قوی به منظور گرفتن تغییرات محیط را نشان می‌دهد. توسعه‌های انجام شده بر مبنای C- میانگین فازی می‌باشند.

پارامترهای مورد استفاده :

C : تعداد کلاس‌ها

N : تعداد اعضا

P : تعداد ویژگی‌های هر عضو

X_i : بردار ویژگی عضو i

V_j : مرکز کلاس j

μij : درجه عضویت عضو I به کلاس j

جدول (۱-۲) : روش‌های خوشه بندی برای داده‌کاوی پویا

کلاس‌های ایستا	کلاس‌های پویا
اعضای ایستا	خوشه بندی مجموعه تغییرات بردارهای ویژگی
اعضای پویا	خوشه بندی مجموعه تغییرات مسیرهای ویژگی

۱.۹.۲ اختصاص اعضای پویا به کلاس‌های ایستا

در بسیاری از کاربردهای دنیای واقعی مقدار ویژگی‌های فعلی به منظور توضیح علت اصلی پدیده کافی نیستند. در چنین وضعیت‌هایی ممکن است علاقه مند باشید این مقادیر را در طول زمان به عنوان ویژگی‌هایی که به صورت مسیر می باشند، توسعه دهید.

۱.۱.۹.۲ روش به کار بردن مسیرها در خوشه بندی

مسیرهای گنجانده شده در سیستم داده‌کاوی می توانند در طول مرحله تبدیل فرایند KDD انجام شوند و سپس روش خوشه بندی کلاسیک اجرا شود یا مستقیماً در همان محدوده روش‌ها اصلاح شوند. تبدیل مسیرهای اصلی به مقادیر واقعی بردارهای ویژگی با حفظ اطلاعات پویای این مسیرها انجام می شود. به خاطر داشته باشید که غالباً کاهش مسیرهای اولیه به مجموعه‌ای از مقادیر واقعی در حالت کلی به معنی از بین رفتن اطلاعات است. FFCM که در ادامه شرح داده می‌شود با تمام مسیرهایی که از مقیاس فاصله اصلاح شده استفاده می‌کند، کار می‌کند.

۲.۱.۹.۲ تشریح عملکرد C- میانگین فازی

به منظور خوشه بندی اعضای پویا، به یک مقیاس فاصله بین دو بردار نیاز است در جایی که هر مولفه یک مسیر (تابع) است بجای عدد واقعی. FFCM یک الگوریتم خوشه بندی فازی است که در آن فاصله مربوطه بر اساس شباهت بین دو مسیر و با استفاده از تابع عضویت تعیین شده است. FFCM یک C- میانگین فازی استاندارد عمومی است. در هر تکرار FCM از فرمول زیر به منظور محاسبه درجه عضویت عضو i در کلاس j استفاده می شود:

$$\mu_{i,j} = \frac{1}{\sum_{k=1}^c \left(\frac{d(x_i, v_j)}{d(x_i, v_k)} \right)^{\frac{2}{m-1}}} \quad (۸-۲)$$

در اینجا x_i بردار ویژگی عضو i ، v_j و v_k مرکز کلاس k و j ، c تعداد کلاس‌ها، $m \in (1, \infty)$ پارامتر تعیین کننده درجه ایجاد خوشه فازی می باشند.

$D(x,y)$ استفاده شده در فرمول بالا، فاصله بین دو بردار x, y را در فضا تعیین می کند. ایده اصلی FFCM تعمیم محاسبه فاصله بین یک جفت از بردارهای ویژگی با اعداد واقعی به محاسبه فاصله بین یک جفت از همین بردارها با مسیرهایشان است. ایده ارائه شده در اغلب تکنیک های داده کاوی، هر جا که فاصله بین اعضا لازم است، می تواند بکار رود.

FFCM فاصله بین دو عضو را با انجام پنج مرحله زیر تعیین می کند:

۱. یک مجموعه فازی A (تقریباً صفر) با تابع عضویت μ تعریف کنید.
۲. درجه عضویت $\mu(f(x))$ از یک تابع اختیاری برای مجموعه فازی A در هر نقطه محاسبه کنید. این درجه عضویت می تواند به عنوان (pointwise) شباهت تابع f به تابع صفر

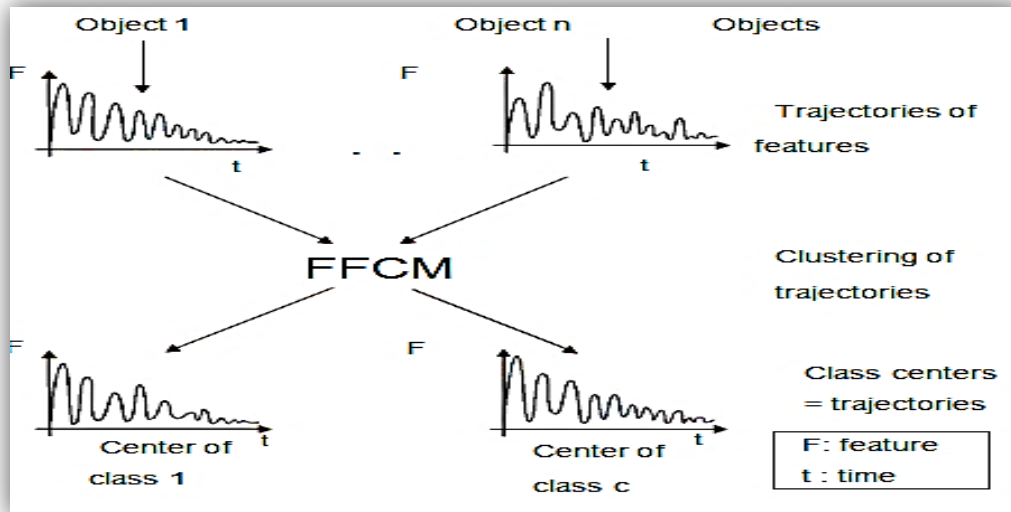
تفسیر شود.

۳. $\mu(f(x))$ تبدیل شده به اعداد واقعی، میزان صفر شدن با بکارگیری تغییرات خاص را نشان می دهد.

۴. میزان شباهتی که در مرحله ۱ تا ۳ تعریف شده، برابر است با نسبت اضافه شده به تابع. به عبارت دیگر $s(f,g)=s(f+h,g+h)$ برای همه توابع f, g, h را نگه می دارد. این اجازه می دهد که شباهت بین یک جفت اختیاری از توابع f و g را توسط تابع افزایشگر $h=g-f$ محاسبه شود و شباهت $f-g$ با تابع شروع تعیین شود.

۵. در آخر، شباهت $s(f,g)$ تبدیل می شود به فاصله $d(f,g)$ با استفاده از: $d(f,g)=1/s(f,g)$

۶. بکارگیری این مقیاس فاصله جدید بین توابع $FFCM$ در تعیین کلاس اعضای پویا در FCM موثر است.



شکل (۲-۸): تشریح عملکرد FFCM

۲.۹.۲ اختصاص اعضای ایستا به کلاس‌های پویا

روشی برای داده‌کاوی پویا بر اساس خوشه بندی فازی که اعضای ایستا را به کلاس‌های پویا اختصاص می‌دهد و یا به عبارت دیگر ترکیب کلاس‌ها را در هر زمان تغییر می‌دهد، ارائه شده است. این کار با یک طبقه بندی معین و مجموعه‌ای از اعضای جدید (اعضایی که بعد از طبقه بندی فعلی ظاهر می‌شوند) شروع می‌شود. زمان میان ایجاد یک طبقه و به روز رسانی آن چرخه نامیده می‌شود. طول چرخه به کاربرد خاص آن بستگی دارد.

۱.۲.۹.۲ نگاه جامع به روش پیشنهاد شده

روش پیشنهاد شده فرض می‌کند که اعضای دیگر قابل شناسایی نیستند. به عبارتی نیازی نیست هر شی‌ای را شناسایی کنید. برای مثال در بخش بندی مشتریان بانک‌ها می‌توانید هر مشتری و فعالیتش را شناسایی کنید ولی در بخش بندی مشتریان سوپر مارکت‌ها، اطلاعات شخصی آن‌ها ندارید که می‌تواند کاربرد این روش برای اعضای غیر قابل شناسایی باشد.

تغییرات ممکن در ترکیب یک خوشه در هر چرخه عبارتند از:

- ایجاد کلاس جدید
- حذف کلاس
- جابجایی کلاس در فضای ویژگی

به منظور نشان دادن تغییر در ترکیب کلاس و پی بردن به تغییرات متناظر پنج مرحله زیر اجرا می‌شود:

۱. شناسایی اعضای که نشان دهنده تغییرات هستند.

ابتدا بررسی کنید که آیا اعضای جدید با خوشه‌های قبلی به خوبی توضیح داده می‌شوند یا نه. یعنی

اعضایی را شناسایی کنید که ترکیب خوشه را تغییر می دهند و برای آن خوشه مناسب نیستند. اگر اعضای زیادی وجود دارند که چنین تغییراتی را نشان می دهند به گام ۲ بروید و در غیر اینصورت به گام ۳ بروید.

۲. تغییر ترکیب هر کلاس را تعیین کنید.

در اینجا تصمیم بگیرید که به منظور اصلاح خوشه ها برای اعضای جدید، کلاس جدید ایجاد کنید و یا فقط جابجایی کلاس کافی است؟ اگر در گام ۱ تعداد اعضای شناسایی شده که نشان دهنده تغییرات هستند زیاد بود، کلاس جدید ایجاد کنید در غیر این صورت آنها را به کلاس های موجود در فضا جابجا کنید.

۳. ترکیب کلاس را تغییر دهید.

مطابق با نتایج گام های ۱ و ۲ تغییرات را ایجاد کنید. (جابجایی (۱) یا ایجاد کلاس (۲))

۳.۱: جابجایی کلاس: مراکز کلاس های موجود را بر اساس اطلاعات آماده شده توسط اعضای جدید به روز رسانی کنید، در حالی که این اعضا نیازی به کلاس جدید ندارند.

۳.۲: ایجاد کلاس: اگر بدانید که باید کلاس جدیدی ایجاد کنید، ابتدا تعداد مناسب کلاس را مشخص کنید سپس C میانگین فازی را با تعداد کلاس های جدید برای داده های در دسترس اجرا کنید.

۴. شناسایی مسیرهای کلاس ها:

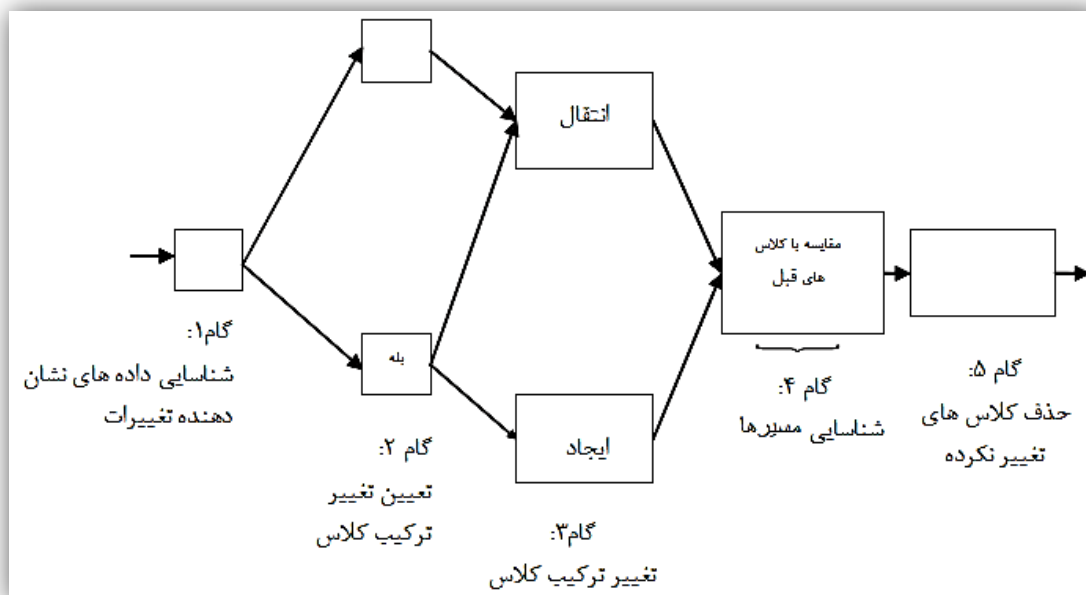
به منظور بررسی وضعیت پذیرش اعضای جدید، مسیر کلاس ها را از چرخه قبل شناسایی کنید؟

کلاس هایی که در طول چندین چرخه عضو جدید دریافت نکرده اند، باید حذف شوند.

۵. حذف کلاس های تغییر نکرده:

طبق نتیجه گام ۴ کلاس هایی که در طی یک دوره پذیرش، عضو جدیدی نپذیرفته اند؛ حذف می

شوند.



شکل (۹-۲): نمای کلی روش خوشه بندی فازی پویا با تغییر ترکیب کلاس ها

۲.۲.۹.۲ شرح جزئیات روش پیشنهاد شده

در طول چرخه m عضو جدید ظاهر شدند. $k = n + 1; \dots; n + m$ نماد m عضو جدید می باشند.

گام ۱: شناسایی عضوهای نشان دهنده تغییرات. هدف از این گام شناسایی عضوهایی است که در خوشه های تعیین شده به خوبی طبقه بندی نشده اند. برای این منظور به فاصله هر زوج از مراکز کلاس های فعلی نیاز دارید. فاصله هر زوج به این صورت است:

$$d(v_i, v_j) \quad \forall i \neq j, i, j \in \{1, \dots, c\} \quad (9-2)$$

علاوه بر این به فاصله بین k عضو جدید و مرکز کلاس v_j از ترکیب کلاس فعلی نیاز دارید. این فاصله به صورت زیر است:

$$\hat{d}_{ik} = \hat{d}(X_k, V_i) \quad i \in \{1, \dots, c\} \quad k \in \{1+n, \dots, n+m\} \quad (10-2)$$

در آخر برای پذیرش m عضو جدید از خوشه های قبلی استفاده کنید.

$$\mu_{i,k} = \text{membership of new object } k \text{ to class } i \quad i \in \{1, \dots, c\} \quad k \in \{1+n, \dots, n+m\}$$

بر اساس اقدامات قبلی c عضو نشان دهنده تغییرات را در خوشه ها شناسایی کردید. در این مرحله می خواهید بدانید که ایجاد کلاس جدید کافی است یا جابجایی آن. از این جهت عضوهایی را که با خوشه های موجود به خوبی طبقه بندی نشده اند و از مرکز کلاس های فعلی خیلی دور هستند، شناسایی کنید. طبق دو شرط زیر این اعضا را تعیین کنید:

$$|\hat{\mu}_{ik} - 1/c| \leq \alpha \quad \forall k \in \{1+n, \dots, n+m\} \quad \forall i \in \{1, \dots, c\} \quad \text{شرط ۱:}$$

مقدار $\alpha \geq 0$. حالت ۱ اعضایی را که درجه عضویت آنها با تعداد کلاس رابطه وارونه دارد، تعیین می کند ($1/c$). این اعضا به صورت مناسب خوشه بندی نمی شوند.

پارامتر آلفا یکی از استراتژی های زیر را برای اقدام تعیین می کند:

- اعضا می توانند ثابت بمانند اگر دانش مربوطه موجود باشد.
- اگر بدانید که طبقه بندی اعضا در چرخه معین درست است، الف را بصورت پویا تعیین کنید که به میزان درستی طبقه بندی اعضا بستگی دارد. برای مثال اگر میزان درستی در چرخه، زیاد باشد، الف باید در چرخه نزدیک به صفر باشد به این منظور که تغییرات کمتری در ترکیب کلاس ها ایجاد شود.

$$\hat{d}_{ik} > \min\{d(v_i, v_j)\} \quad \forall k \in \{1+n, \dots, n+m\} \quad \forall i \notin \{1, \dots, c\} \quad \text{شرط ۲:}$$

این حالت تعیین میکند که آیا k عضو جدید از کلاس‌های فعلی خیلی دور هستند. فرض کنید که این گونه است که فاصله آن‌ها از i بزرگتر از \min فاصله بین دو کلاس i و j است. بر این اساس دو حالت تعریف می‌شود:

$$1/c(X_k) = \begin{cases} 1 & \text{هر دو شرط برای } X_k \text{ برقرار باشد} \\ 0 & \text{در غیر این صورت} \end{cases}$$

به عبارت دیگر $1/c(X_k)$ مقدار ۱ را می‌پذیرد اگر و فقط اگر عضو k در خوشه‌های فعلی به خوبی خوشه بندی نشود.

اگر $\sum_{k=n+1}^{n+m} 1/c(X_k) = 0$ به گام ۳ بروید در غیر این صورت گام ۲ را ادامه دهید.

گام ۲: تعیین تغییرات در ترکیب کلاس‌ها. با توجه به اینکه حداقل یک عضو برای تغییر در ترکیب کلاس طبقه نیاز است، چک کنید که آیا به کلاس جدید نیاز دارید یا جابجایی کلاس موجود کافی است. برای این کار طبق ضابطه زیر عمل کنید:

$$\frac{\sum_{k=n+1}^{n+m} 1/c(X_k)}{m} \geq \beta \quad \text{with a parameter } \beta \quad 0 \leq \beta \leq 1 \quad (11-2)$$

اگر رابطه بین اعضای جدیدی که نمایانگر تغییرات هستند و کل عضوهای جدید (m) در ضابطه بالا صدق کرد، کلاس جدید ایجاد کنید و در غیر این صورت فقط کلاس موجود جابجا شود.

β یکی از استراتژی‌های زیر را برای اقدام تعیین می‌کند:

- این مفهوم می‌تواند ثابت باشد در صورتیکه دانش مربوط به آن موجود باشد.

- اگر طبقه اعضا در چرخه معین به درستی شناخته شود، β می تواند میزان درستی طبقه را تعیین کند.

گام ۳: تغییر در ترکیب کلاس. بر اساس نتایج گام ۲ ممکن است کلاس جدیدی تشکیل شود و یا یکی از کلاس های موجود جابجا خواهد شود.

۳-۱: جابجایی کلاس: دو گزینه اساسی برای جابجایی کلاس وجود دارد:

- اجرای الگوریتم خوشه بندی اصولی (C میانگین فازی) با اعضا قبلی و جدید (بدون تغییر در تعداد کلاس)

- تعیین مرکز کلاس نمایانگر اعضای جدید و ترکیب آن ها با مراکز کلاس های قبلی.

در حالت دوم انتقال کلاس، مراکز کلاس های موجود را با مراکز نمایانگر اعضای جدید وابسته به همان کلاس ترکیب می کند. برای این منظور تابع نماینده k عضو جدید برای کلاس i را تعریف کنید:

$$1_{C_i}(X_k) = \begin{cases} 1 & \text{object } k \text{ is assigned to class } i, \\ 0 & \text{else.} \end{cases}$$

هر عضو را به کلاسی اختصاص دهید که بیشترین مقدار عضویت در آن کلاس را دارا است. برای هر عضو i، مرکز کلاسی که فقط نشان دهنده اعضای جدید باشد تعیین کنید.

$$v_i^* = \frac{\sum_{k=n+1}^{n+m} (1 - 1/c(X_k)) (\hat{\mu}_{ik})^m X_k}{\sum_{k=n+1}^{n+m} (1 - 1/c(X_k)) (\hat{\mu}_{ik})^m} \quad 1 \leq i \leq c \quad (12-2)$$

با ترکیب این (مراکز عضوهای جدید) با مراکز قبلی، مرکز کلاس جدید تعیین کنید:

$$\hat{v}_i = (1 - \lambda_i) v_i + \lambda_i v_i^*, \quad (13-2)$$

در اینجا وزن λ_i نسبت تخصیص اعضای جدید را به کلاس i نشان می دهد.

$$\lambda_i = \frac{\sum_{k=n+1}^{n+m} [1_{C_i}(\mathbf{x}_k) \cdot (1 - 1_{IC}(\mathbf{x}_k)) \cdot \hat{\mu}_{ik}]}{\sum_{j=1}^n (1_{C_i}(\mathbf{x}_j) \cdot \mu_{ij}) + \sum_{k=n+1}^{n+m} [1_{C_i}(\mathbf{x}_k) \cdot (1 - 1_{IC}(\mathbf{x}_k)) \cdot \hat{\mu}_{ik}]}.$$

(۱۴-۲)

۲-۳: ایجاد کلاس: اگر تشخیص دادید که به یک یا چند کلاس جدید نیاز دارید، در صورتی که تعداد عضوهای جدیدی که در کلاس های موجود بطور مناسب تخصیص داده نشدند زیاد باشد، ابتدا تعداد کلاس های مناسب را برای آن ها تعیین کنید. برای این منظور مفهوم ارائه شده توسط لی و موکایدونو^{۳۷} را که دوام ساختار نامیده می شود، بکار ببرید. بر اساس این مفهوم پیشنهاد شده که با اطلاع از یک بخش، به راحتی نتیجه درست را تخمین بزنید. تابع خطا $L(c)$ از میان مجموع مجذور خطاهای (WGSS) برای یک راه حل خوشه بندی معین تعیین می شود.

$$L(c) = \sum_{k=1}^N \sum_{i=1}^c u_{ik} * d_{ik}^2,$$

(۱۵-۲)

C : تعداد کلاس ها

N : تعداد اعضا

u_{ik} : درجه عضویت عضو k به کلاس i

d_{ik} : فاصله بین عضو k و مرکز کلاس i

بر اساس تابع خطا، تعداد کلاس ها به صورت زیر تعیین می شود:

$S(c) = \text{structure strength} =$

$\alpha^*(\text{effectiveness of classification}) + (1 - \alpha)^*(\text{accuracy of classification}) =$

$\alpha^* \log(N/c) + (1 - \alpha)^* \log(L(1)/L(c)).$

³⁷.Li and Mukaidono

موکایدنو اندازه گیری اثربخش را با استفاده از $\log(N/C)$ ارائه کرد. به عبارت دیگر طبقه بندی با کلاس های کمتر موثرتر است. او اندازه گیری با $\log(L(1)/L(c))$ را نیز پیشنهاد کرد، به عبارت دیگر طبقه بندی با کلاس های بیشتر درست تر است. $L(1)$ واریانس مجموعه داده های درست و الفا وزن بین صحت و اثربخشی است. مولف پیشنهاد کرده که برای یک تخمین بدون غرض (اثربخشی و صحت مقادیر یکسان داشته باشند) مقدار $\alpha = 0.5$ باشد و مقدار C ماکزیمم $S(c)$ فرض شود که تعداد کلاس ها کافی باشد. با استفاده از تعداد مناسب کلاس های جدید برای همه اعضا، الگوریتم خوشه بندی پایه (در اینجا C میانگین فازی) را به منظوره تعیین بهترین C^{new} کلاس های نشان دهنده همه اعضا، اجرا کنید.

گام ۴: شناسایی مسیرها: ایجاد کلاس و یا انتقال در این مرحله ضرورت دارد. بررسی کنید که آیا کلاسی برای حذف وجود دارد؟ جهت آمادگی برای مرحله حذف کردن (مرحله ۵)، تغییر هر کلاس را در طول چرخه قبلی بر اساس مسیرهایش و C_t^i برای کلاس i در چرخه t شناسایی کنید. در این جا دو گزینه زیر را دارید:

کلاس i در چرخه t ایجاد شده است. در این مورد شماره آن را $C_t^i = 1$ بگذارید.

کلاس i نتیجه انتقال قطعی کلاس j در چرخه $t-1$ است. در این مورد $C_t^i = C_{t-1}^j + 1$ است.

گام ۵: حذف کلاس های بدون تغییر: حذف یک کلاس با روش زیر تعیین می شود:

" کلاسی که طول یک دوره معین عضو جدید دریافت نکرده باشد، باید حذف شود"، که دوره معین باید تعریف شود.

در گام ۵ برای هر کلاس i در چرخه t شماره C_t^i تعریف می شود. به عبارت دیگر تعداد چرخه های آن فعال شده است. ماکزیمم تعداد چرخه های یک کلاس که می توانند فعال باشند را بدون اعضای جدید دریافت شده تعیین کنید (اینجا T چرخه). اگر کلاسی برای T چرخه عضو جدید دریافت نکند، حذف

خواهد شد. در روش پیشنهاد شده برای همه کلاس‌ها از مقدار T یکسان استفاده می‌شود که توسط کاربر تعیین می‌شود. در برخی از کاربردها می‌توان از مقادیر مختلف T برای کلاس‌های موجود استفاده کرد.

باید گفت که حذف یک کلاس به معنی این نیست که از این کلاس کاملاً صرف نظر شده است. زمانی که یک کلاس حذف می‌شود، اطلاعات مربوط به آن در حافظه جداگانه‌ای نگهداری می‌شود. اگر در چرخه بعدی در گام ایجاد کلاس، کلاس جدیدی ایجاد شود، آن خیلی به کلاس حذف شده قبلی شبیه است (ویر^{۳۸}، ۲۰۰۷).

۱۰.۲ منطق فازی

در سال ۱۹۶۵ در دانشگاه کالیفرنیا، برکلی پرفسور لطفی زاده تئوری مجموعه‌های فازی و منطق فازی^{۳۹} را مطرح کرد، وی برای اولین بار با معرفی نظریه مجموعه‌های فازی مقدمات مدل سازی اطلاعات نادقیق و استدلال تقریبی با معادله‌های ریاضی را فراهم نمود که در نوع خود تحولی عظیم در ریاضیات و منطق کلاسیک بوجود آورد. ایده نظریه مجموعه‌های فازی با این عبارت توسط پرفسور لطفی زاده مطرح شد: "ما نیازمند یک نوع دیگری از ریاضیات هستیم تا بتوانیم ابهامات و عدم دقت رویدادها را مدل سازی نماییم، مدلی که متفاوت از نظریه احتمالات است." لذا نظریه فازی برای بیان و تشریح عدم قطعیت و عدم دقت در رویدادها به کار می‌رود که براساس منطق چند ارزشی بوجود آمده است (شوندی، ۱۳۸۵).

۱.۱۰.۲ مجموعه‌های فازی

³⁸ . Webre

³⁹ .Fuzzy lojic

در منطق کلاسیک عضویت در یک مجموعه به صورت صفر و یک در نظر گرفته می‌شود؛ بدین صورت که در صورتی که عضوی در یک مجموعه وجود داشته باشد با ۱ و در غیر این صورت با ۰ نشان داده می‌شود. و در حقیقت درجه عضویت تابعی است که برد آن عضو مجموعه $\{0,1\}$ می‌باشد. اما از طرف دیگر در منطق فازی، مفهوم درجه عضویت در یک مجموعه به بازه $[0, 1]$ گسترش می‌یابد. مفهوم منطق فازی از آن جهت مورد توجه قرار می‌گیرد که در جهان واقع نیز بسیاری از استدلال‌ها و دلایل بشری، جنبه عدم قطعیت و تقریبی دارد.

تعریف مجموعه فازی: یک مجموعه فازی روی یک مجموعه مبدا X مجموعه‌ای از جفت‌های

$$A = \{\mu_A(x)/x : x \in X, \mu_A(x) \in [0, 1] \in R\}$$

به صورتی که $\mu_A(x)$ تابع درجه عضویت عضو فازی x مجموعه A نامیده می‌شود. تابع درجه عضویت می‌تواند هر یک از مقادیر حقیقی بین ۰ و ۱ را بپذیرد.

$\mu_A(x) = 0$: بیانگر این است که x قطعاً به مجموعه فازی A تعلق ندارد.

$\mu_A(x) = 1$: بیانگر این است که x قطعاً به مجموعه فازی A تعلق دارد.

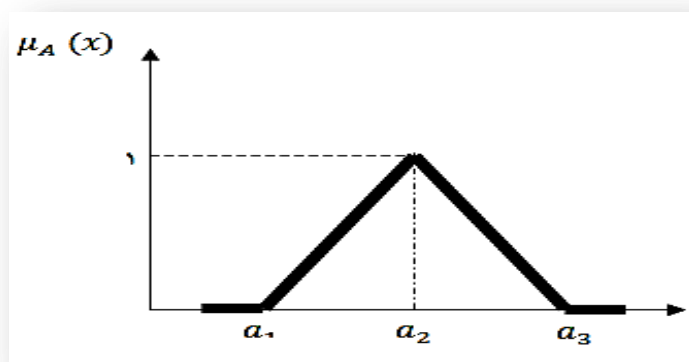
در ادامه مهمترین خصوصیات منطق فازی آمده است:

- در منطق فازی، استدلال دقیق یا منطق معمولی حالت خاصی از استدلال تقریبی است.
- هر سیستم منطقی قابل تبدیل به منطق فازی است.
- در منطق فازی، دانش به عنوان مجموعه‌ای از محدودیت‌های فازی یا انعطاف پذیر روی متغیرها در نظر گرفته می‌شود.
- استنتاج به عنوان فرآیند انتشار این محدودیت‌ها در نظر گرفته می‌شود.

• در منطق فازی تمام مسائل دارای راه حلی هستند که درجه مطلوبیت (امکان) را نشان می‌دهد.

برای توابع عضویت انتخاب‌های متفاوتی وجود دارد که بسته به کاربرد مد نظر می‌توان یکی از آنها را انتخاب کرد. در یک تقسیم بندی کلی که توسط زاده ارائه شد می‌توان توابع فازی را به دو دسته خطی و غیر خطی (منحنی) تقسیم بندی کرد. توابع مثلثی، یکه، L، گاما، دوزنقه، S، گاوسی، شبه نمایی، از جمله معروف‌ترین توابعی هستند که برای مدل کردن درجه عضویت در مجموعه های فازی برای کاربردهای متفاوت مورد استفاده قرار گرفته‌اند (جرج^{۴۰}، ۱۹۹۵).

شکل زیر یک عدد فازی را نشان می‌دهد که با این تعریف مطابق است. لازم به ذکر است که اعداد فازی به وسیله ی حروف برجسته ی A, B, C, \dots و یا به صورت $\tilde{a}, \tilde{b}, \tilde{c}, \dots$ و تابع عضویت آن ها به وسیله ی $\mu_A(x), \mu_B(x), \mu_C(x), \dots$ نمایش داده می شوند.



شکل (۲-۱۰): عدد فازی مثلثی مثبت

عدد فازی مثلثی مثبت (TFN) A ، با یک سه تایی به صورت $A = (a_1, a_2, a_3)$ نشان داده می‌شود و

تابع عضویت آن به صورت زیر است:

⁴⁰. George

$$\mu_A(x) = \begin{cases} 0, & x < a_1 \\ \frac{x - a_1}{a_2 - a_1}, & a_1 \leq x \leq a_2 \\ 1, & x = a_2 \\ \frac{x - a_3}{a_2 - a_3}, & a_2 \leq x \leq a_3 \\ 0, & x > a_3 \end{cases}$$

علاوه بر اعداد فازی مثلثی، اعداد فازی ذوزنقه‌ای نیز از عمومی‌ترین و پرکاربردترین اعداد فازی می‌باشند. در بخش‌های آتی نیز خواهیم دید که نویسنده برای ایجاد تابع فازی NPS از هر دوی این اعداد و توابع آن‌ها استفاده نموده است. عدد فازی ذوزنقه‌ای A بصورت $A = (a_1, b_1, b_2, a_2)$ ، نشان داده می‌شود. در این حالت بازه ی تکیه گاه برابر $[b_1, b_2]$ بوده و بخش هم سطح در سطح $\alpha = 1$ ، روی محور X دارای تصویر $[b_1, b_2]$ است. تابع عضویت اعداد فازی ذوزنقه‌ای در زیر آورده شده است.

$$\mu_A(x) = \begin{cases} 0, & x < a_1 \\ \frac{x - a_1}{b_1 - a_1}, & a_1 \leq x \leq b_1 \\ 1, & b_1 \leq x \leq b_2 \\ \frac{x - a_2}{b_2 - a_2}, & b_2 \leq x \leq a_2 \\ 0, & x > a_2 \end{cases}$$

همچنین با فرض در اختیار داشتن دو عدد فازی \tilde{m} و \tilde{n} ، روابط ریاضی ذیل برای آن‌ها برقرار خواهد

بود.

$$\tilde{m} (+) \tilde{n} = (m_1 + n_1, m_2 + n_2, m_3 + n_3)$$

$$\tilde{m} (-) \tilde{n} = (m_1 - n_1, m_2 - n_2, m_3 - n_3)$$

$$\tilde{m} \otimes \tilde{n} = (m_1 \cdot n_1, m_2 \cdot n_2, m_3 \cdot n_3)$$

$$\tilde{m} (\div) \tilde{n} = \left(\frac{m_1}{n_1}, \frac{m_2}{n_2}, \frac{m_3}{n_3} \right)$$

۲.۱۰.۲ میانگین فازی

۱.۲.۱۰.۲ فرمول میانگین مثلثی

چنان چه n عدد مثلثی بصورت $\mathbf{A}_i = (a_1^i, a_M^i, a_2^i)$ در اختیار داشته باشیم، \mathbf{A}_{ave} برابر است با:

$$\begin{aligned} A_{ave} &= \frac{A_1 + \dots + A_n}{n} \\ &= \frac{(a_1^1, a_M^1, a_2^1) + \dots + (a_1^n, a_M^n, a_2^n)}{n} \\ &= \frac{(\sum_{i=1}^n a_1^i, \sum_{i=1}^n a_M^i, \sum_{i=1}^n a_2^i)}{n} \\ &= \left(\frac{1}{n} \sum_{i=1}^n a_1^i, \frac{1}{n} \sum_{i=1}^n a_M^i, \frac{1}{n} \sum_{i=1}^n a_2^i \right) \quad A_{ave} = (m_1, m_M, m_2) \end{aligned}$$

۲.۲.۱۰.۲ فرمول میانگین دوزنقه‌ای

اگر $\mathbf{A}_i = (a_1^i, b_1^i, b_2^i, a_2^i)$ و $i = 1, \dots, n$ اعداد دوزنقه‌ای باشند، پس:

$$\begin{aligned} A_{ave} &= (m_1, m_{M1}, m_{M2}, m_2) \\ &= \frac{(a_1^1, b_1^1, b_2^1, a_2^1) + \dots + (a_1^n, b_1^n, b_2^n, a_2^n)}{n} \\ &= \frac{(\sum_{i=1}^n a_1^i, \sum_{i=1}^n b_1^i, \sum_{i=1}^n b_2^i, \sum_{i=1}^n a_2^i)}{n} \end{aligned}$$

۳.۱۰.۲ فازی زدایی از میانگین فازی

انبوهش تعریف شده به وسیله ی میانگین دوزنقه‌ای یا مثلثی، به طور غالب بایستی به وسیله ی یک عدد معین نشان داده شود که بهترین میانگین متناظر را معرفی می‌کند، این عملیات فازی زدایی نامیده می‌شود. اگر داشته باشیم: $A_{ave} = (m_1, m_M, m_2)$ ، به نظر می‌رسد برای نیل به هدف فازی زدایی، عدد m_M درون بازه تکیه‌گاه $[m_1, m_2]$ از A_{ave} را انتخاب نماییم. m_M بالاترین درجه ی عضویت ۱ را در

A_{ave} دارد. به عبارت دیگر A_{ave} ماکزیمم درجه را در m_M حاصل می کند، که آن را ماکزیمم کننده ارزش می نامند.

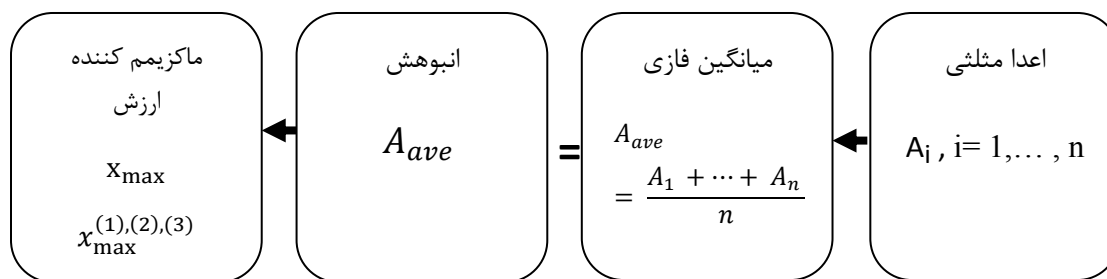
در این جا سه گزینه انتخابی برای فازی زدایی $(m_1, m_M, m_2) = A_{ave}$ معرفی می شود که در اصل رابطه های مربوط به میانگین آماری هستند:

$$x_{max}^{(1)} = \frac{m_1 + m_M + m_2}{3}$$

$$x_{max}^{(2)} = \frac{m_1 + 2m_M + m_2}{4}$$

$$x_{max}^{(3)} = \frac{m_1 + 4m_M + m_2}{6}$$

بر خلاف حالت قبل، رابطه های میانگین، سهم m_1 و m_2 را در نظر می گیرند، اما به m_M اوزان مختلف اختصاص می دهند. اگر عدد مثلثی A_{ave} نزدیک به عدد مثلثی مرکزی باشد، به این معنی است که m_M تقریباً در وسط $[m_1, m_2]$ قرار می گیرد و $x_{max} = m_M$. به طور معمول، اعداد میانگین مثلثی، در کاربردها، به شکل مرکزی حاصل می شوند. با این حال کارشناسان گاه با شرایطی روبرو می شوند که باید از قضاوت خود، برای انتخاب ماکزیمم کننده ارزش استفاده کنند. شکل زیر، روش فازی زدایی را به صورت یک نمودار بلوکی نمایش می دهد.



فازی زدایی میانگین فازی

۱۱.۲ بخش بندی و تحلیل رفتار مشتریان

دانشمندان روش‌های متنوعی را در زمینه خوشه بندی معرفی نموده اند. این روش‌ها مبتنی بر مفروضات خاص و بر اساس نوع داده‌ها، شکل خوشه‌ها، فاصله داده‌ها و غیره دارای عملکرد متفاوتی می‌باشند.

غضنفری و همکاران (۱۳۸۹) با استفاده از الگوریتم k - میانگین و k - میانگین فازی مشتریان صادراتی میوه‌های خوراکی را خوشه بندی نموده و نتایج این خوشه بندی را از لحاظ کیفیت بر اساس سه معیار مختلف سنجیده اند. مرتضوی و همکاران (۱۳۹۰) با استفاده از تحلیل خوشه‌ای بر مبنای ملاک بیزی بازار گوشی تلفن همراه را بر مبنای مزایای مورد انتظار مشتریان به سه خوشه تقسیم بندی نموده و آن‌ها را مورد مقایسه قرار دادند. ابطحی و اکبرزاده (۱۳۹۱) الگوریتم خوشه‌بندی فازی را با استفاده از جمعیت پویا و الگوهای مبتنی بر جمعیت بهینه‌سازی کردند.

خان بابایی و زین العابدینی (۱۳۹۲) با ارائه مدلی بر مبنای فرایند استاندارد کریسپ در داده‌کاوی اقدام به شناسایی و تحلیل رفتار مشتریان خدمات بانکداری الکترونیکی کردند و نشان دادند که مدل پیشنهادی می‌تواند مشتریان را بر اساس رفتار آن‌ها شناسایی و تحلیل نماید و به بخش بندی و طبقه بندی آن‌ها بپردازد. بررسی تکنیک‌های بکار گرفته شده در تحقیقات دانشمندان نشان می‌دهد علی‌رغم وجود محیط رقابتی و تغییر پذیری در تحلیل رفتار مشتریان فرض تغییری پذیری داده‌ها در دامنه زمانی نادیده گرفته شده‌اند. خیری و احمدی (۱۳۹۳) عوامل موثر بر قصد تغییر رفتار مشتریان و پیامدهای آن را مورد بررسی قرار دادند و با استفاده از روش تحلیل چند متغیره و الگویابی معادلات ساختاری روابط بین متغیرها را مشخص نمودند.

لیندر و همکاران^{۴۱} (۲۰۰۴) برای بخش‌بندی مشتریان سه روش شبکه عصبی مصنوعی، درخت تصمیم و رگرسیون لجستیک را به کار گرفته که بر اساس مقایسه آن‌ها نشان داده‌اند که روش مبتنی بر شبکه‌های عصبی ابزار کارآمدتری برای مدل‌سازی رفتار مشتریان با داده‌های پیچیده و نمونه‌های کوچک است. کیم و آن^{۴۲} (۲۰۰۸) با استفاده از سه تکنیک شبکه SOM^{۴۳}، الگوریتم K- میانگین و روش تلفیقی K- میانگین و الگوریتم ژنتیک مشتریان یک سایت اینترنتی ارائه دهنده خدمات رژیم را در ۵ بخش تقسیم‌بندی کرده و برتری نسبی روش تلفیقی و شبکه SOM را تایید کردند.

چانگ و شن^{۴۴} (۲۰۰۸) در پژوهش خود با به کارگیری تکنیک خوشه‌بندی بر مبنای مدل RFM موزون و محاسبه مقدار ارزش دوره زندگی مشتری برای هر خوشه و سپس امتیازدهی خوشه‌ها بخش بندی جدیدی از مشتریان به منظور بازاریابی هدف و استراتژی‌های فروش متقاطع در حوزه صنعت فروشگاه‌های بزرگ انجام دادند. وو و همکاران^{۴۵} (۲۰۰۹) با استفاده از تکنیک خوشه‌بندی K- میانگین ارزش مشتریان یک شرکت ساخت تجهیزات صنعتی را مورد تحلیل قرار دادند و پیشنهاداتی برای برنامه‌های ترفیع مناسب با هر بخش از مشتریان ارائه کردند.

ناندا و همکاران^{۴۶} (۲۰۱۰) الگوریتم خوشه‌بندی K- میانگین را برای بخش‌بندی مشتریان سهام در خوشه‌های متنوع به منظور ساخت یک مدل مدیریت پورتفوی با دارا بودن کمترین میزان ریسک مورد استفاده قرار داده‌اند. لی و همکاران^{۴۷} (۲۰۱۱) با استفاده از روش خوشه‌بندی دو مرحله‌ای به تحلیل ویژگی‌های مشتریان یک کارخانه بافندگی پرداختند. مبنای تحلیل خوشه‌ای در این پژوهش مدل توسعه یافته RFM بوده است. در این پژوهش مشتریان با روش k- میانگین به پنج خوشه تقسیم و تحلیل

⁴¹ . Linder, Geier, Kolliker

⁴² . Kim, Ahn

⁴³ .Self Organization Map

⁴⁴ .Chuang ,Shen

⁴⁵ . Wu, chang.Lo

⁴⁶ .Nanda, Mahanty, Tiwari

⁴⁷ . Li, Dai, Tseng, Ean

ویژگی‌های هر خوشه با استفاده از مدل RFM انجام گرفته است. چانفی و ژی یی^{۴۸} (۲۰۱۳) در پژوهش خود الگوریتم خوشه بندی k - میانگین را بر اساس مراکز اولیه و مقدار k بهبود دادند. الگوریتم بهبود یافته از داده‌های نویزی موجود در پایگاه داده اجتناب می‌کند و یا آن‌ها را کاهش می‌دهد و نتیجه خوشه بندی نهایی را واقعی‌تر و موثرتر می‌کند.

سرت و همکاران^{۴۹} (۲۰۱۴) رویکرد جدیدی برای فرایند درک رفتار پویای مشتریان بر اساس ترکیب تکنیک خوشه‌بندی و روش دنباله‌کاوی ارائه کرده‌اند که بر اساس آن مسیر رفتار مشتریان در پایگاه داده کشف می‌شود.

بکارگیری الگوریتم خوشه بندی پویا^{۵۰} (DFCM) در تحلیل تغییرپذیری رفتار مشتریان که در این تحقیق ارائه شده بدیع و جدید می‌باشد.

⁴⁸ . Chunfei, Zhiyi

⁴⁹ . Seret , Seppe, Broucke , Baesens, Vanthienen

⁵⁰ .Dynamic Fuzzy C-means

فصل سوم

روش تحقیق

۱.۳ مقدمه

این پژوهش پس از تعریف مسئله و سوالات تحقیق و بررسی جامع ادبیات موضوع پیرامون داده‌کاوی و خوشه‌بندی و همچنین روش‌های داده‌کاوی پویا سعی دارد با ارائه طرح تحقیق مناسب و اجرای الگوریتم نوین خوشه‌بندی جامعه مشتریان گوشی تلفن همراه سامسونگ را طی دوره چهار ساله معین بخش بندی نماید. در ادامه با بررسی مقایسه‌ای روش ایستا و پویا بهترین روش خوشه‌بندی تعیین خواهد شد. برنامه نویسی و اجرای الگوریتم در محیط نرم افزار MATLAB 2008a اجرا شده است.

در این فصل نوع پژوهش و روش جمع‌آوری داده‌ها و جامعه آماری و روش تحلیل داده‌ها به تفصیل مورد بررسی قرار خواهد گرفت.

۲.۳ نوع پژوهش

پژوهش‌های علمی را از دو جهت می‌توان دسته‌بندی نمود:

۱. **هدف تحقیق:** تحقیقات بر اساس هدف تحقیق به دو دسته تحقیقات بنیادی و کاربردی تقسیم می‌شوند.

● پژوهش بنیادی: پژوهشی است که به کشف ماهیت اشیا پدیده‌ها و روابط بین متغیرها، اصول، قوانین و ساخت یا آزمایش تئوری‌ها و نظریه‌ها می‌پردازد و به توسعه مرزهای دانش رشته علمی کمک می‌نماید.

هدف اصلی این نوع پژوهش تبیین روابط بین پدیده‌ها، آزمون نظریه‌ها و افزودن به دانش موجود در یک زمینه خاص است.

- پژوهش کاربردی: پژوهشی است که با استفاده از نتایج تحقیقات بنیادی به منظور بهبود و به کمال رساندن رفتارها، روش‌ها، ابزارها، وسایل، تولیدات، ساختارها و الگوهای مورد استفاده جوامع انسانی انجام می‌شود. هدف تحقیق کاربردی توسعه دانش کاربردی در یک زمینه خاص است. تحقیق حاضر با توجه هدف از نوع تحقیقات کاربردی است.

۲. بر اساس نحوه گردآوری داده‌ها پژوهش‌ها به دو دسته توصیفی و آزمایشی تقسیم می‌شوند:

- پژوهش‌های توصیفی شامل پنج دسته می‌باشد: پیمایشی، همبستگی، پس‌رویدادی، اقدام پژوهشی، بررسی موردی.

- پژوهش آزمایشی که به دو دسته آزمایشی و نیمه آزمایشی تقسیم می‌شوند. تحقیق حاضر از نظر نحوه گردآوری داده‌ها جز تحقیقات توصیفی - پیمایشی می‌باشد.

۳.۳ روش گردآوری داده‌ها

یکی از اصلی‌ترین بخش‌های هر کار پژوهشی جمع‌آوری اطلاعات می‌باشد. چنان‌چه این کار به شکل منظم و صحیح صورت پذیرد کار تجزیه و تحلیل و نتیجه‌گیری از داده‌ها با سرعت و دقت خوبی انجام خواهد شد.

امروزه به دلیل وجود ابزارهای مختلف برای جمع‌آوری داده‌ها و پیشرفت قابل قبول تکنولوژی پایگاه داده، حجم انبوهی از اطلاعات در انبار داده‌های مختلف ذخیره شده است. در این پژوهش آن‌چه که به عنوان داده مورد بررسی و تجزیه و تحلیل قرار گرفته می‌گیرد به صورت مجموعه داده از پیش تهیه شده می‌باشد. با توجه به حجم بالای داده‌های ذخیره شده، جهت استخراج اطلاعات و کشف الگوهای مفید (با حداقل دخالت کاربر) از این مجموعه از فرآیند داده‌کاوی استفاده خواهد شد. در داده‌کاوی از تحلیل اکتشافی داده‌ها استفاده می‌شود که در آن بر کشف اطلاعات نهفته و ناشناخته از درون انبوهی از داده‌ها

تاکید می‌شود. وجود حجمی انبوه از داده‌ها پیش‌فرض داده‌کاوی است. هر چه حجم داده‌ها بیشتر و روابط میان آنها پیچیده تر باشد داده‌کاوی اهمیت بیشتری پیدا می‌کند.

۴.۳ جامعه و نمونه آماری

جامعه مورد نظر این پژوهش کلیه افرادی هستند که طی چهار سال گذشته حداقل یک نوع از گوشی‌های تلفن همراه برند سامسونگ را از یکی از ۱۵۵ نمایندگی فعال این شرکت در شهر تهران خریداری کرده‌اند.

اطلاعات مربوط به مشتریان و خرید انجام گرفته توسط آن‌ها در پایگاه داده این نمایندگی‌ها ذخیره و به عنوان داده‌های مورد نیاز پژوهش مورد استفاده قرار گرفته‌اند. در این پژوهش اطلاعات مربوط به خرید سایر محصولات این شرکت از پایگاه داده حذف شده و تنها اطلاعات ۳۰۰۰ نفر از مشتریان که تلفن همراه خریداری نموده‌اند مورد قبول واقع شده است.

از آن‌جا که اصلی‌ترین مرحله تصمیم‌گیری در یک سیستم خوشه‌بندی انتخاب ویژگی‌های مناسب است، لذا هر چه تعداد ویژگی‌های انتخاب شده مناسب‌تر باشد، خوشه‌بندی دقیق‌تر خواهد بود. بر این اساس در این پژوهش به هر مشتری یک بردار ویژگی نسبت داده شده که این بردار پارامترهای لازم برای مشخص‌سازی هر مشتری و وضعیت خرید وی را دارا می‌باشد. در ادامه ویژگی‌های مورد نیاز برای خوشه‌بندی ارائه خواهند شد. جدول زیر این ویژگی را نشان می‌دهد.

جدول (۱-۳) : ویژگی های مشتریان و محصولات

ردیف	ویژگی	توضیحات
۱	Sex	جنسیت مشتریان (مرد - زن)
۲	Age	سن مشتریان
۳	Income	درآمد مشتریان
۴	RAM	حافظه موقت (512 M,1 G,2 G)
۵	Price	قیمت گوشی تلفن همراه
۶	Sim	تعداد سیم کارت (تک سیم کارت - دو سیم کارت)
۷	Camera	کیفیت دوربین (بدون دوربین - ۵ MP و کمتر - ۸ MP و بیشتر)
۸	CPU	پردازشگر تلفن همراه (تک هسته ای - دو هسته ای - چهار هسته ای)

کلیه ویژگی‌هایی که در رابطه با خوشه بندی در این پژوهش مطرح شده اند به صورت بصری و تجربی انتخاب شده‌اند. ویژگی‌های Sex,Ram,Sim,Camera,CPU به صورت کد و سایر ویژگی‌ها با مقادیر واقعی در پایگاه داده وجود دارند.

داده‌های ثبت شده در پایگاه داده مورد استفاده به صورت زیر می‌باشند:

جدول (۳-۲): پایگاه داده

Customer	Sex	Age	Income / 1000	RAM	Price / 1000	Sim	Camera	CPU
۱	۱	۱۹	۸۴۰	۱	۱۳۰	۲	۱	۱
۲	۱	۳۴	۱۵۰۰	۱	۶۳۰	۱	۲	۲
۳	۱	۲۸	۲۴۵۰	۱	۱۲۰۰	۲	۳	۱
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
۲۹۹۸	۲	۲۶	۹۸۹	۱	۳۶۰	۱	۲	۱
۲۹۹۹	۲	۳۲	۳۶۵۸	۱	۱۸۵۲	۲	۳	۳
۳۰۰۰	۱	۴۷	۲۱۰۰	۱	۹۵۶	۲	۳	۳

اطلاعات مربوط به مشتریان در پنج بازه زمانی به ترتیب در پایگاه داده ثبت شده‌اند.

- دوره اول: ۱۰۰۰ نفر
- دوره دوم: ۵۰۰ نفر
- دوره سوم: ۴۰۰ نفر
- دوره چهارم: ۶۰۰ نفر
- دوره پنجم: ۵۰۰ نفر

۵.۳ روش‌های تجزیه و تحلیل داده‌ها

مبنای اصلی این پژوهش تکنیک خوشه بندی فازی پویای ارائه شده توسط ریچارد وبر می‌باشد. در طول پژوهش همواره سعی شده که از اصول آن فاصله نگرفته و تنها در جهت درک بهتر نتایج حاصل از اجرای تکنیک، تمامی داده‌ها با مبنای ثابت نیز خوشه بندی شده و نتایج حاصل از هر دو روش مورد

مقایسه قرار گرفتند. به منظور خوشه بندی مشتریان در هر دو روش از الگوریتم FCM استفاده شده است که بر اساس آن هر مشتری به خوشه‌ای تعلق می‌گیرد که با اعضای آن خوشه همگن تر باشد. نحوه تحلیل داده‌ها در روش پویا به شرح زیر می‌باشد:

۱. داده‌های مربوط به سه ماه اول در سه کلاس (C=3) خوشه بندی می‌شوند و مراکز خوشه‌ها و تعداد اعضای هر خوشه تعیین می‌شوند.

۲. داده‌های دو ماه دوم به مجموعه قبلی اضافه شده و الگوریتم FCM دوباره اجرا خواهد شد و مراکز جدید خوشه‌ها و تعداد اعضای آنها مشخص می‌شوند. (چرخه ۱)

با اضافه شدن داده‌های جدید در هر چرخه ممکن است هر یک از تغییرات زیر در ترکیب کلاس‌ها مشاهده شود:

- ایجاد کلاس جدید
- حذف کلاس
- انتقال کلاس در فضای ویژگی‌ها

به منظور شناسایی داده‌های ایجاد کننده تغییر در ترکیب کلاس‌ها در هر چرخه موارد زیر برای داده‌های جدید بررسی خواهد شد.

$$|\hat{\mu}_{ik} - 1/C| \leq \alpha \quad \forall k \in \{1+n, \dots, n+m\} \quad \forall i \in \{1, \dots, c\} \quad (3-1)$$

μ_{ik} : درجه عضویت عضو i ام به خوشه k ام (خوشه تعلق گرفته)

در این پژوهش $C=3$ و $\alpha = 0.05$ در نظر گرفته خواهد شد.

$$\hat{d}_{ik} > \min\{d(v_i, v_j)\} \quad \forall k \in \{1+n, \dots, n+m\} \quad \forall i \notin \{1, \dots, c\} \quad (3-2)$$

\hat{d}_{ik} : فاصله عضو i ام از مرکز خوشه k ام

$d(v_i, v_j)$: فاصله مراکز خوشه های i ام و j ام از یکدیگر

بر اساس دو شرط فوق حالت های زیر تعریف می شوند:

$$1/c(x_k) = \begin{cases} 1 & \text{هر دو شرط صادق باشند} \\ 0 & \text{حداقل یکی از شرط ها برقرار نباشد} \end{cases}$$

داده هایی که در هر دو شرط صدق می کنند از مراکز کلاس های فعلی دور بوده و به خوبی خوشه بندی نخواهند شد و ترکیب کلاس را تغییر خواهند داد. برای تعیین نوع تغییر از ضابطه زیر استفاده خواهد شد:

$$\frac{\sum_{k=n+1}^{n+m} 1/c(x_k)}{m} \geq \beta \quad (3-3)$$

M : کل داده های جدید

در این پژوهش $\beta = 0.05$ در نظر گرفته می شود.

برای داده هایی که در این ضابطه صدق می کنند خوشه جدید تشکیل می شود، در غیر این صورت فقط مراکز کلاس ها منتقل شده اند.

۳. داده های دو ماه سوم را به مجموعه داده های قبلی اضافه می کنیم و مجددا الگوریتم FCM اجرا

خواهد شد و فرایند گام ۲ تکرار می شود. (چرخه ۲)

در این مرحله روند تغییرات کلاس ها را در گام های قبلی بررسی خواهیم کرد چنان چه تغییرات کلاس ها طی T چرخه متوالی کمتر از ۰.۵ باشد، آن کلاس به دلیل ثبات رفتار حذف خواهد شد.

در این پژوهش $T=2$ خواهد بود.

۴. برای سه ماه چهارم (چرخه ۳) و دو ماه پنجم (چرخه ۴) نیز به همین ترتیب الگوریتم اجرا خواهد شد تا همه داده‌ها مورد بررسی قرار بگیرند و تغییرات کلاس‌ها ثابت گردد. پس از اجرای الگوریتم خوشه بندی پویا، داده‌ها را به صورت زیر با مبنای ثابت خوشه بندی خواهیم کرد:

۱. داده‌های مربوط به سه ماه اول در سه کلاس خوشه بندی می‌شوند و مراکز خوشه‌ها و تعداد اعضای هر خوشه تعیین می‌شوند.

۲. در این مرحله فاصله اقلیدسی (d) بین داده‌های داده‌های جدید (دو ماه دوم) را از مراکز خوشه‌های گام قبلی (مراکز کلاس‌های سه ماه اول) تعیین کرده و این داده‌ها را بر اساس حداقل فاصله به خوشه‌ها اختصاص می‌دهیم.

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2} \quad (۴-۳)$$

۳. الگوریتم FCM برای مجموع داده‌ها گام‌های قبل اجرا خواهد شد و مراکز کلاس جدید تعیین می‌شود.

۴. داده‌های دو ماه سوم اضافه شده و این فرایند از گام دوم تکرار می‌شود. به این ترتیب در هر مرحله فاصله داده‌های جدید از مراکز کلاس‌های قبلی تعیین و بر اساس حداقل فاصله به هر کلاس عضو جدید تعلق خواهد گرفت.

پس از انجام خوشه بندی به هر دو روش و تعیین تغییرات مراکز کلاس‌ها و تعداد اعضای هر کلاس، در هر مرحله مجموع فاصله اقلیدسی اعضای هر کلاس را از مرکز همان کلاس (مجموع خطا) تعیین خواهیم کرد.

بدیهی است که هر چه مجموع فاصله اقلیدسی اعضا از مراکز کلاس‌ها کمتر باشد، خوشه بندی دقیق‌تر و مناسب‌تر خواهد بود. بررسی این شاخص به شناسایی تکنیک بهتر برای خوشه بندی و تعیین میزان بهبود خطا در روش برتر نسبت به روش دیگر کمک خواهد نمود.

در پایان نیز مجموع داده‌های مربوط به پنج بازه زمانی ذکر شده بصورت انباشته (۳۰۰۰ نفر) با دو تکنیک FCM و K-means خوشه بندی خواهد شد و نتایج آن‌ها مورد تجزیه و تحلیل قرار خواهد گرفت.

فصل پہلے ہمارے

تجزیہ و تحلیل دادہ

۴-۱ مقدمه

در این بخش نتایج مربوط به تجزیه و تحلیل داده‌ها ارائه می‌گردد. در بخش اول تحلیل داده‌ها به صورت دوره‌ای انجام خواهد شد و هر دوره شامل چهار قسمت خواهد بود. قسمت اول تحلیل اطلاعات مربوط به خوشه بندی پویا و قسمت دوم مربوط به خوشه بندی با مبنای ثابت می‌باشد. در قسمت سوم هر دوره مجموع خطا هر دو روش ارائه می‌شود و در قسمت چهارم شکل خوشه‌ها براساس سه ویژگی (درآمد- سن- قیمت) نمایش داده می‌شود.

در بخش دوم تحلیل داده به صورت تجمعی انجام خواهد شد و تمامی داده‌ها به صورت فازی و قطعی خوشه بندی می‌شوند. در هر بخش داده‌ها بر اساس نرم افزار MATLAB 2008a تحلیل خواهد شد و به سوالات مطرح شده در تحقیق پاسخ خواهیم داد.

۴-۲ تجزیه و تحلیل دوره‌ای داده‌ها

قبل از تجزیه و تحلیل داده‌ها یادآوری می‌کنیم این بخش شامل پنج دوره زمانی می‌باشد که در فصل سوم به آن اشاره شد.

۱-۲-۴ تحلیل داده‌های دوره اول

جدول (۱-۴): مراکز خوشه‌های پویا برای $C=3$

Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number	Changes
Dynamic FCM	1	1.5002	30.9728	37.0457	1.5155	14.7559	1.4072	2.0606	1.6859	238	-
	2	1.4812	37.9604	14.3379	1.5215	12.8121	1.4560	2.0719	1.6220	305	-
	3	1.4515	26.2990	14.3771	1.5317	16.8555	1.4339	2.0855	1.6921	457	-
	Total =1000									1000	

جدول (۲-۴): مراکز خوشه‌های ایستا برای $C=3$

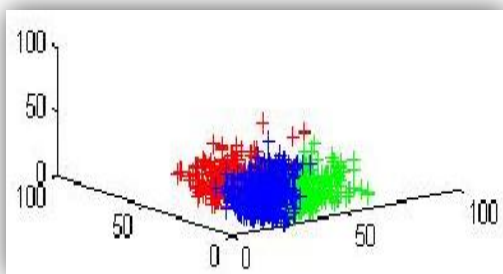
Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number	New Object
Static FCM	1	1.4515	26.2990	14.3771	1.5317	16.8555	1.4339	2.0855	1.6921	457	-
	2	1.4812	37.9604	14.3379	1.5215	12.8121	1.4560	2.0719	1.6220	305	-
	3	1.5002	30.9728	37.0457	1.5155	14.7559	1.4072	2.0606	1.6859	238	-
	Total =1000									1000	

در جدول (۱-۴) اطلاعات خوشه‌های تشکیل شده با روش پویا و در جدول (۲-۴) مشخصات خوشه‌های ایجاد شده با مبنای ثابت برای داده‌های سه ماه اول ارائه شده‌اند.

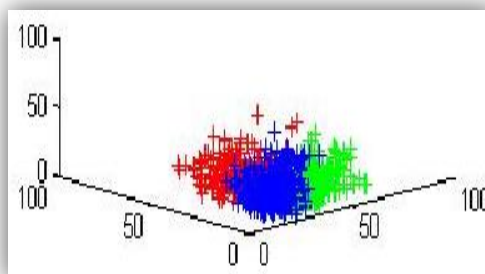
همان‌طور که مشخص است مشخصات خوشه‌های تشکیل شده در این مرحله برای هر دو روش یکسان می‌باشند و تنها تفاوت مربوط به اولویت تشکیل خوشه‌ها در هر مرحله می‌باشد. در این مرحله مشتریان در همه خوشه‌ها تمایل به خرید گوشی تلفن همراه تک سیم کارته با کیفیت دوربین ۵MP و پردازشگر دو هسته‌ای و رم ۲G دارند. تفاوت خوشه‌ها ناشی از سطح درآمد افراد و قیمت گوشی‌ها و جنسیت و سن افراد می‌باشد.

جدول (۳-۴) : مجموع خطا

Errore	Cluster Number	Daynmic FCM	Static FCM
$\sum_{i=1}^{1000} d_i$	1	4675.3785	3387.6249
	2	3483.8958	3483.8958
	3	3387.6249	4675.3785
Total		11546.8992	11546.8992



شکل (۱-۴) : خوشه بندی پویا



شکل (۲-۴) : خوشه بندی ایستا

با توجه به یکسان بودن خوشه‌ها، مجموع خطا در این مرحله برای هر دو روش یکسان می‌باشد.

۲-۲-۴ تحلیل داده‌های دوره اول و دوم

جدول (۴-۴): مراکز خوشه‌های پویا در چرخه ۱

Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number	Changes
Dynamic FCM	1	1.4663	38.3448	14.4271	1.5053	13.6339	1.5076	2.2595	1.7250	443	Movement
	2	1.4543	26.3370	14.3835	1.5202	16.1139	1.4566	2.2585	1.7513	695	Movement
	3	1.4987	31.2136	37.2366	1.4787	14.6536	1.4562	2.2361	1.7326	362	Movement
	Total =1500									1500	

جدول (۴-۵): مراکز خوشه‌های ایستا در چرخه ۱

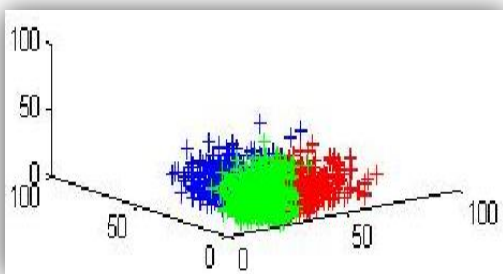
Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number	New Object
Static FCM	1	1.4515	26.2990	14.3771	1.5317	16.8555	1.4339	2.0855	1.6921	457	214
	2	1.4812	37.9604	14.3379	1.5215	12.8121	1.4560	2.0719	1.6220	305	161
	3	1.5002	30.9728	37.0457	1.5155	14.7559	1.4072	2.0606	1.6859	238	125
	Total =1500									1000	500

داده‌های مربوط به دو ماه دوم که شامل ۵۰۰ نفر بودند به داده‌های دوره اول اضافه شد و نتایج حاصل از هر روش به صورت زیر مشخص شد.

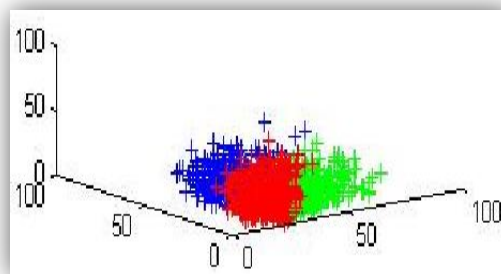
همان‌طور که مشخص است در این مرحله نتایج حاصل از دو روش با هم متفاوت می‌باشد. در روش ایستا (جدول (۴-۵)) ذائقه افراد هر خوشه ثابت باقی می‌ماند و فقط اعضای خوشه‌ها بیشتر شده‌اند در صورتی که ذائقه افراد در روش پویا (جدول (۴-۶)) دستخوش تغییر شده است.

جدول (۶-۴): مجموع خطا در چرخه ۱

Erore	Cluster Number	Daynmic FCM	Static FCM
$\sum_{i=1}^{1500} d_i$	1	288.2899	6986.2473
	2	453.8562	5303.5861
	3	236.3521	5117.3195
Total		978.4982	17407.1529



شکل (۳-۴): خوشه بندی پویا چرخه ۱



شکل (۴-۴): خوشه بندی ایستا چرخه ۱

جدول (۶-۴) نشان می‌دهد با اجرای الگوریتم‌ها در این مرحله مجموع خطا برای روش پویا بهبود پیدا کرد ولی برای روش ایستا مجموع خطا افزایش یافت.

۳-۲-۴ تحلیل داده‌های سه دوره

جدول (۷-۴): مراکز خوشه‌های پویا در چرخه ۲

Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number	Changes
Dynamic FCM	1	1.4589	26.4998	14.1651	1.5230	15.2070	1.4818	2.3064	1.7767	892	Movement
	2	1.5002	30.7654	36.9554	1.4966	14.8157	1.4729	2.3012	1.7672	453	Movement
	3	1.4593	38.9805	14.6854	1.5069	14.8026	1.5116	2.3462	1.7611	555	Movement
	Total =1900									1900	

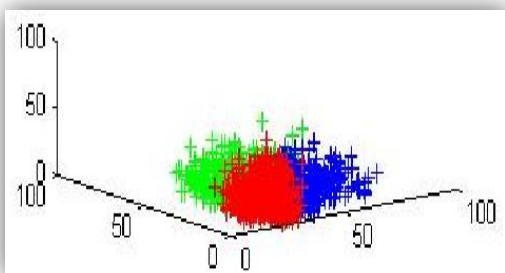
جدول (۸-۴): مراکز خوشه‌های ایستا در چرخه ۲

Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number	New Object
Static FCM	1	1.4987	31.2136	37.2366	1.4787	14.6536	1.4562	2.2361	1.7326	362	89
	2	1.4543	26.3370	14.3835	1.5202	16.1139	1.4566	2.2585	1.7513	695	179
	3	1.4663	38.3448	14.4271	1.5053	13.6339	1.5076	2.2595	1.7250	443	132
	Total =1900									1500	400

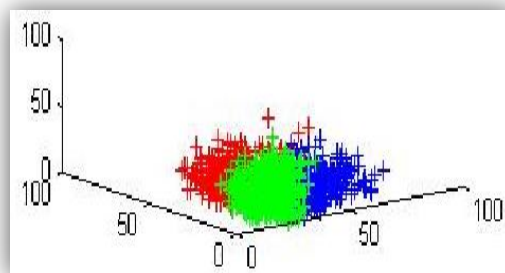
براساس نتایج جدول (۷-۴) مشخص است که اضافه شدن داده‌های دوره سوم تغییراتی چندانی در ذائقه جامعه ایجاد نکرده است. به عبارت دیگر ذائقه افراد جدید بسیار شبیه به ذائقه افراد قبلی جامعه بوده است.

جدول (۹-۴): مجموع خطا در چرخه ۲

Erore	Cluster Number	Daynmic FCM	Static FCM
$\sum_{i=1}^{1900} d_i$	1	589.0895	6353.1048
	2	298.1007	9062.8837
	3	359.3161	6681.1317
Total		1246.5063	22097.1202



شکل (۵-۴): خوشه بندی پویا چرخه ۲



شکل (۶-۴): خوشه بندی ایستا چرخه

در این مرحله نیز مجموع خطا در روش پویا نسبت به روش ایستا بسیار کمتر است.

۴-۲-۴ تحلیل داده‌های چهار دوره

جدول (۱۰-۴): مراکز خوشه های پویا در چرخه ۲

Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number	Changes
Dynamic FCM	1	1.4798	31.0623	13.6213	1.5117	15.0066	1.5183	2.2572	1.8191	1399	Movement
	2	1.5151	32.3354	30.9576	1.4983	15.4377	1.5048	2.2920	1.8230	522	Movement
	3	-	-	-	-	-	-	-	-	-579	Elimination
	Total =2500									1921	

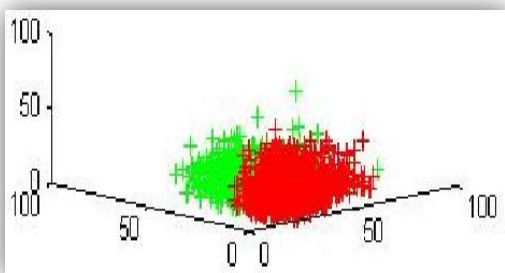
جدول (۱۱-۴): مراکز خوشه های ایستا در چرخه ۳

Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number	New Object
Static FCM	1	1.4593	38.9805	14.6854	1.5069	14.8026	1.5116	2.3462	1.7611	555	178
	2	1.4589	26.4998	14.1651	1.5230	15.2070	1.4818	2.3064	1.7767	892	295
	3	1.5002	30.7654	36.9554	1.4966	14.8157	1.4729	2.3012	1.7672	453	127
	Total =2500									1900	600

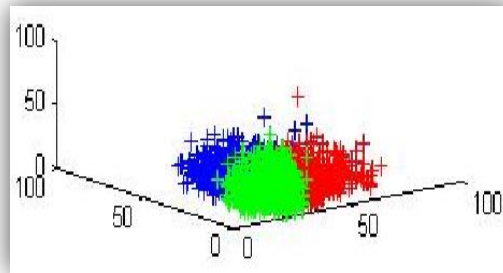
نتایج جدول (۴-۱۰) نشان می‌دهد که در ترکیب خوشه‌ها تغییرات عمده‌ای رخ داده است. مشخص است که تعدادی از داده‌ها رفتار ثابتی داشته و در بررسی تغییرات رفتاری نقشی نداشته‌اند و خوشه‌ای که این داده‌ها در بر داشته است از جامعه حذف شده است. اما در روش ایستا همه داده‌ها حفظ شده و نتایج آن با روش پویا بسیار متفاوت می‌باشد.

جدول (۴-۱۲): مجموع خطا در چرخه ۳

Errore	Cluster Number	Daynmic FCM	Static FCM
$\sum_{i=1}^{2500} d_i$	1	1060.7757	8660.8828
	2	388.7183	12133.1449
	3	0	8080.6512
Total		1449.4895	28874.6789



شکل (۴-۷): خوشه بندی پویا چرخه ۳



شکل (۴-۸): خوشه بندی ایستا چرخه ۳

با توجه به حذف داده‌های ثابت در روش پویا مجموع خطا در این روش بسیار کمتر از روش ایستا می

باشد.

۴-۲-۵ تحلیل داده‌های پنج دوره

جدول (۴-۱۳): مراکز خوشه پویا در چرخه ۴

Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number	Changes
Dynamic FCM	1	1.4969	30.7151	36.7393	1.4433	14.9259	1.5444	2.1853	1.9576	567	Movement
	2	1.4816	26.6656	14.1981	1.4607	14.7833	1.5539	2.1811	1.9539	1141	Movement
	3	1.4782	39.2972	14.7496	1.4532	15.8762	1.5662	2.2235	1.9651	713	Creation
	Total =3000									2421	

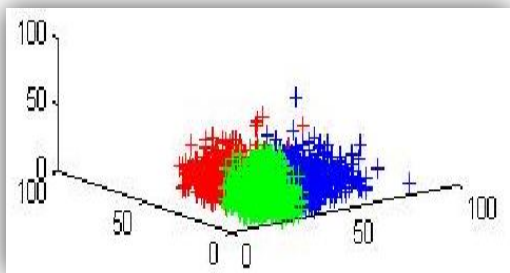
جدول (۴-۱۴): مراکز خوشه ایستا در چرخه ۴

Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number	New Object
Static FCM	1	1.4627	34.4179	30.9859	1.4183	16.2534	1.5868	2.2788	2.0490	763	147
	2	1.4794	30.7407	13.8649	1.4563	15.0555	1.5606	2.1971	1.6917	1737	353
	Total =3000									2500	500

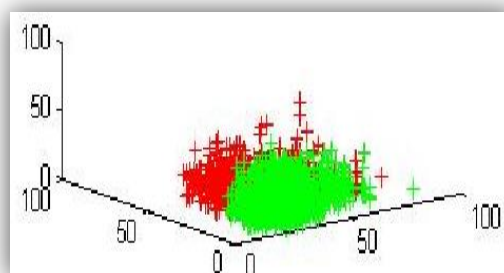
جدول (۴-۱۳) نشان می‌دهد که داده‌های جدید این مرحله با مشخصات خوشه‌های قبلی سازگار نبوده و از این جهت خوشه جدید تشکیل شده است. اما در روش ایستا نتیجه نهایی با دو خوشه نشان داده شده است.

جدول (۴-۱۵): مجموع خطا در چرخه ۴

Erore	Cluster Number	Daynmic FCM	Static FCM
$\sum_{i=1}^{3000} d_i$	1	374.1187	18353.0470
	2	759.1125	32712.1236
	3	459.5045	0
Total		1592.7357	51065.283



شکل (۴-۹): خوشه بندی پویا چرخه ۴



شکل (۴-۱۰): خوشه بندی ایستا در چرخه ۴

در جدول (۴-۱۵) مجموع خطا برای ۳۰۰۰ داده طی پنج مرحله مشخص شده است. بدیهی است که روش پویا عملکرد بسیار بهتری نسبت به روش ایستا دارد.

۳-۴ تجزیه و تحلیل تجمعی داده‌ها

۱-۳-۴ تحلیل قطعی

جدول (۴-۱۶): مراکز خوشه های قطعی

Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number
K-means	1	1.4935	30.8742	39.2057	1.4387	14.6846	1.5403	2.1677	1.9532	620
	2	1.4848	46.0738	15.4143	1.4559	14.5088	1.5875	2.3082	1.9727	623
	3	1.4764	26.6936	13.7703	1.4549	9.2738	1.5533	2.1814	1.9641	976
	4	1.4930	28.3291	14.3482	1.4686	24.7018	1.5749	2.1869	1.9718	781
	Total =3000									3000

۲-۳-۴ تحلیل فازی

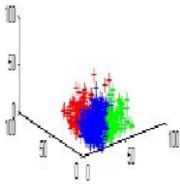
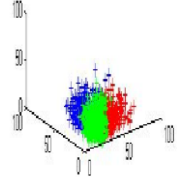
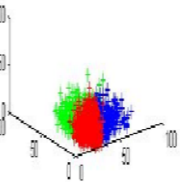
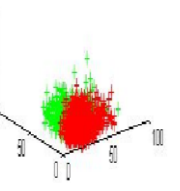
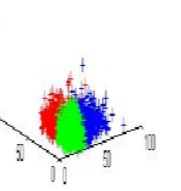





جدول (۱۷-۴): مراکز خوشه های فازی

Parameter	Cluster Number	Sex	Age	Income	RAM	Price	Sim	Camera	CPU	Object Number
FCM	1	1.4985	28.0156	14.7776	1.4541	20.7368	1.5537	2.1691	1.9528	826
	2	1.4737	27.8126	14.2228	1.4454	9.5699	1.5650	2.1969	1.9451	866
	3	1.4831	42.4731	15.4393	1.4461	16.0263	1.5739	2.2200	1.9910	685
	4	1.4997	30.6853	39.3020	1.4475	14.7569	1.5393	2.1677	1.9493	620
	Total =3000									3000

نتایج جدول های (۱۶-۴) و (۱۷-۴) نشان می دهد که بین روش های فازی و قطعی تفاوت چندانی وجود ندارد و تنها یکی از خوشه ها در هر دو جدول تا حدودی با هم اختلاف دارند. شبیه بودن نتایج روش فازی و قطعی به دلیل ماهیت داده ها در این پژوهش می باشد. داده های این تحقیق با درجه عضویت بسیار زیادی به هر خوشه تعلق گرفته اند. با توجه به این موضوع می توان گفت احتمال تعلق گرفتن یک داده به خوشه خاص بسیار زیاد و تا حدی قطعی می باشد. از این جهت نتایج این دو روش بسیار به هم نزدیک می باشد.

۳-۵ تجزیه و تحلیل تغییرات رفتاری

جدول (۴-۱۸) : نتایج تحلیل رفتار مشتریان

بازه های زمانی	خوشه بندی اولیه	چرخه اول	چرخه دوم	چرخه سوم	چرخه چهارم									
	۱۰۰۰	۱۵۰۰	۱۹۰۰	۲۵۰۰ با حذف ۵۷۹	۳۰۰۰									
نمودار														
نمودار														
	سبز	آبی	قرمز	سبز	آبی	قرمز	سبز	آبی	قرمز	آبی	قرمز	سبز	آبی	قرمز
جنسیت	مرد	خانم	مرد	مرد	مرد	مرد	مرد	خانم	مرد	خانم	مرد	مرد	مرد	مرد
سن	۲۶	۳۰	۳۸	۲۶	۳۱	۳۸	۲۶	۳۱	۳۹	۳۲	۳۱	۲۶	۳۰	۳۹
درآمد میلیون	۱	۲.۶	۱	۱	۲.۷	۱	۱	۲.۶	۱	۲.۲	۰.۹۷	۱	۲.۶	۱
قیمت میلیون	۰.۷۸	۰.۷	۰.۶	۰.۷۷	۰.۷	۰.۶۵	۰.۷۲	۰.۷	۰.۷	۰.۷۳	۰.۷	۰.۷	۰.۷	۰.۷۵
دوربین MP	۵	۵	۵	۸	۸	۸	۸	۸	۸	۸	۸	۸	۸	۸
پردازشگر	دو	دو	دو	دو	دو	دو	دو	دو	دو	دو	دو	چهار	چهار	چهار
تعداد سیم	۱	۱	۱	۱	۱	۲	۱	۱	۲	۲	۲	۲	۲	۲
RAM	۱	۲	۱	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲

خوشه بندی اولیه: در این مرحله جامعه مشتریان شامل ۱۰۰۰ نفر می باشد که با نظر محقق در سه خوشه جای گرفته‌اند. در هر سه خوشه مشخص است که مشتریان تمایل به خرید تلفن همراه تک سیم کارته و دارای دوربین ۵ مگا پیکسلی و پردازشگر دو هسته‌ای می باشند. بازه قیمتی برای تهیه این نوع از تلفن‌های همراه بین ۶۰۰ هزار تا ۷۸۰ هزار تومان می باشد.

چرخه اول: با اضافه شدن ۵۰۰ مشتری جدید در این مرحله سلیقه مشتریان در جهت تهیه تلفن همراه با دوربین قدرتمندتر (۸ مگا پیکسلی) و دارای RAM بالاتر تغییر کرده است. همچنین مشتریان تمایل دارند که در بازه قیمتی ۶۵۰-۷۲۰ هزار تومان خرید خود را انجام دهند.

چرخه دوم: در این دوره اطلاعات خرید ۴۰۰ مشتری دیگر به دیگر اطلاعات اضافه شده است که تغییری در رفتار خرید افراد به وجود نیاورده است. عدم تغییر رفتار می‌تواند دلایل مختلفی داشته باشد مثلاً این که سلیقه افراد جدید با سلیقه مشتریان قبلی به هم شباهت داشته و یا این که در طی این دوره زمانی محصول متفاوت و جدیدی وارد بازار نشده است و مشتریان از محصولات قبلی انتخاب نموده‌اند.

چرخه سوم: در این بازه زمانی اطلاعات خرید ۶۰۰ مشتری جدید به اطلاعات قبلی اضافه شده است. در این مرحله تعداد خوشه‌ها کاهش یافته است. حذف یک خوشه به دلیل ثبات رفتار خرید ۵۷۹ نفر از مشتریان می‌باشد که طی دو بازه زمانی قبلی تغییر رفتار نداده‌اند. این ثبات در رفتار به دلایل متفاوتی می‌تواند اتفاق بیفتد. از جمله این عوامل محدودیت توان خرید افراد است که منجر به انتخاب محصول خاص می‌شود و یا میانگین سنی افراد بالا بوده که می‌تواند یکی از دلایل عدم تغییر ذائقه افراد باشد و دلایل بسیار دیگری که می‌توان از آن‌ها نام برد.

چرخه چهارم: در آخرین مرحله با اضافه شدن اطلاعات ۵۰۰ مشتری جدید دوباره یک خوشه جدید تشکیل می‌شود. اضافه شدن یک خوشه جدید نشان می‌دهد که سلیقه برخی از مشتریان با سلیقه مشتریان دو خوشه قبلی متفاوت بوده و در گروه مجزا قرار می‌گیرد.

فصل پنجم

نیچگیری و پیشهادات

۵-۱ مقدمه

یکی از مهم‌ترین بخش‌های هر پژوهش علمی بیان مهم‌ترین یافته‌ها و دستاوردهای آن می باشد که باید مورد توجه قرار گیرد. نتایج از دو بخش اصلی یافته‌ها و پیشنهادات تشکیل می‌شود که حاصل تمام تلاش محقق را در طول انجام تحقیق در خود جای می‌دهد. چنانچه نتایج با دقت تبیین نشوند بسیاری از تلاش‌های صورت گرفته در اجرای تحقیق فاقد ارزش شده و یا از بین می‌روند. لذا توجه به این بخش از پژوهش از اهمیت بسیاری برخوردار است.

۵-۲ بحث و نتیجه گیری

رفتار خرید مشتریان تابعی از ویژگی‌های فردی و ویژگی‌های محصول خریداری شده است که در فرایند زمان تغییر می‌کند. مرور ادبیات نشان می‌دهد که یکی از اشتباهات رایج در تحلیل رفتار مشتریان فرض ایستا بودن داده‌های جمع آوری شده می‌باشد. اما از آنجا که رفتار مشتریان در طی زمان تغییر می‌کند لذا تحلیل حاصله و تصمیم سازی بازاریابی بایستی مبتنی بر رفتارهای متغیر باشد و تشخیص جهت غالب تغییرات می‌تواند روند تغییرات آتی بازار را نشان دهد.

در تشخیص رفتار مشتریان و طبقه بندی آنها تمرکز غالب مطالعات دانشمندان بر روی تکنیک‌های داده‌کاوی بوده است که با بکارگیری آنها خوشه بندی جدیدی از مشتریان ایجاد می‌شود. روش‌های تفکیکی، روش‌های سلسله مراتبی و الگوریتم‌های فرا ابتکاری و تکنیک RFM و به عنوان مهم‌ترین تکنیک‌ها در این حوزه می‌باشند. از معایب این تکنیک‌ها این است که در فرض پویا بودن اشیا و پویا بودن خوشه ناتوان می‌باشند. اما محیط واقعی بازاریابی و خوشه بندی رفتار مشتریان با فرض متغیر بودن اشیا و خوشه هم خوانی بیشتری دارند .

در این مقاله با الگوریتم پیشنهادی C- میانگین فازی پویا برای شناسایی و تحلیل رفتار مشتریان نشان داده شده است که با حذف رفتارهای ثابت و تحلیل بر اساس رفتارهای متغیر دسته بندی واقعی تری از رفتار مشتریان ایجاد شده است.

خوشه بندی مشتریان گوشی همراه با ۵ بازه زمانی (بازه اول: ۱۰۰۰ مشتری، بازه دوم: ۵۰۰ مشتری، بازه سوم: ۴۰۰ مشتری، بازه چهارم: ۶۰۰ مشتری، بازه پنجم: ۵۰۰ مشتری) و با ۸ ویژگی (جنسیت، سن، درآمد، قیمت گوشی، تعداد سیم کارت، کیفیت دوربین، پردازشگر، RAM) با دو روش پیشنهادی و روش رایج صورت گرفته است. نتایج حاصله بر اساس دو رویکرد نشان می دهد که مجموع انحرافات اشیاء نسبت به مرکز خوشه ها در روش پیشنهادی کمتر از روش رایج می باشد.

بعلاوه تحلیل فرایند بازه ای نشان می دهد که بخش بندی و رفتار مشتریان در تغییرات فرایند زمانی نسبت به ویژگی های کیفیت دوربین، پردازشگر، تعداد سیم کارت و RAM تغییرات معناداری را داشته و تغییرات ویژگی های سنی و جنسیت و درآمد و قیمت گوشی چندان معنادار نمی باشد. این تغییرات معنادار تعداد خوشه و جایگاه فضائی آن ها را تغییر داده است، بطوری که در بازه سوم تعداد خوشه ۳ تا و در بازه چهارم تعداد خوشه ها ۲ تا بوده است. این تغییرات بیشتر حاصل تغییر و همگرا شدن در ذائقه و رفتار مشتریان در ویژگی سن مشتری و تعداد سیم کارت می باشد. از طرفی وجود ویژگی هایی که تغییرات آن ها معنادار نبوده است نشان می دهد که بخش بندی بازار به ویژگی های زیادی وابسته نبوده و می توان ویژگی ها را کاهش داد و یا ویژگی های دیگری را جایگزین کرد.

۳-۵ محدودیت های تحقیق

محدویت ها از جمله عواملی هستند که در فرایند هر تحقیقی وجود دارند و حرکت به سوی هدف را دچار کندی و چالش می نمایند. یکی از محدودیت های این تحقیق کمبود منابع دست اول در زمینه

مفاهیم مورد استفاده در این پژوهش بود. عدم دسترسی به یک پایگاه داده جامع و مناسب جهت تحقیق مهم‌ترین محدودیت این تحقیق به شمار می‌آید. همچنین به دلیل پیچیدگی الگوریتم پیشنهادی و عدم دسترسی به منابع فارسی و یا منابع لاتین جامع، جهت برنامه نویسی الگوریتم نیاز به دانش و تجربه بالا در زمینه برنامه نویسی بود که این موضوع روند تحقیق را با کندی مواجه کرد.

۴-۵ پیشنهادات کاربردی - مدیریتی

از آنجا که این الگوریتم بدون فرض ایستا بودن داده‌ها در پایگاه داده اجرا می‌شود، می‌تواند تغییر پذیری رفتار افراد را در فرایند زمانی به درستی نشان داده و در تصمیم‌گیری‌های بازاریابی مدیران شرکت مورد استفاده قرار گیرد. بر اساس تحلیل نتایج به مدیران شرکت پیشنهاد می‌شود که در برنامه-ریزی‌های آتی شرکت جهت توسعه محصولات جدید به عامل قیمت توجه ویژه‌ای داشته باشند، زیرا تحلیل‌ها نشان می‌دهد که همه گروه‌های مشتریان نسبت به قیمت حساس بوده و بازه تغییرات قیمت در طول زمان و با افزایش قابلیت‌های گوشی همراه تغییر چندانی نداشته است. بررسی نتایج نشان می‌دهد که چهار ویژگی کیفیت دوربین، پردازشگر، تعداد سیم کارت و RAM بسیار مورد توجه مشتریان بوده و تغییرات در این زمینه می‌تواند تغییرات در سهم بازار را برای شرکت به دنبال داشته باشد. بنابراین توجه به این ویژگی‌ها و تقویت آن‌ها جز نقاط قوت شرکت محسوب می‌شود.

همچنین پیشنهاد می‌شود که برای شناخت دقیق نیازها و ترجیحات مشتریان و ارائه پاسخ متناسب با محیط مشتری روش بخش بندی پویا جایگزین روش های ایستا گردد که هم زمینه تصمیم‌گیری پویا برای توسعه محصول جدید را نیز فراهم نماید و هم هزینه‌های بازاریابی و بخش بندی و تحقیق و توسعه را کاهش دهد.

روش پیشنهادی با ارائه تصویر مناسبی از بازار و با قدرت برآوردکنندگی بالا ابزار مناسبی برای شناسایی بازار هدف و تعیین برنامه‌های بازاریابی و تدوین استراتژی‌های آینده می‌باشد. از این رو به

مدیران پیشنهاد می‌شود که استفاده از این روش را در برنامه های تحلیلی خود جای داده تا در همه برنامه های بازاریابی با کمترین هزینه و صرف وقت با تغییرات محیط و رفتار مشتریان روبرو شوند.

۵-۵ پیشنهاد تحقیقات آتی

از آنجا که بخش بندی و تحلیل رفتار مشتریان در تمام عمر شرکتها به طور مستمر وجود دارد، به طراحان و برنامه نویسان پیشنهاد می‌شود که الگوریتم پیشنهادی را در قالب یک برنامه کاربردی تهیه نمایند که شرکتها با سهولت و انعطاف بیشتری از مزایای این الگوریتم بهره ببرند. همچنین به سایر محققین پیشنهاد می‌شود که در پژوهشهای آتی سایر متغیرها و یا سایر محصولات برای تحلیل رفتار مشتریان مورد استفاده قرار دهند و نتایج را باهم مقایسه نمایند.

تئوری راف به عنوان یکی از تکنیک های استخراج الگوهای رفتاری محسوب شده و در این روش پایگاه داده ها ثابت فرض می‌شود لذا جهت مطالعات آتی پیشنهاد می‌شود تا سایر محققان با بکارگیری الگوریتم تئوری راف مبتنی بر فرض متغیر بودن داده‌ها در بازه زمانی بتوانند قوانین کاربردی و واقعی-تری را استخراج نمایند.

سود


```

datatez=[sex age income edu price sim camera cpu];
initialdata=datatez;
num=[1000;500;400;600;500];numm=num(1,1);sumsum=sum(num);

```

```

[A,c]=kmeans(datatez(:,1:8),4)

```

```

y1=0;y2=0;y3=0;y4=0;

```

```

for i=1:sumsum

```

```

    if A(i,1)==1

```

```

        y1=y1+1;

```

```

    elseif A(i,1)==2

```

```

        y2=y2+1;

```

```

    elseif A(i,1)==3

```

```

        y3=y3+1;

```

```

    elseif A(i,1)==4

```

```

        y4=y4+1;

```

```

    end

```

```

end

```

```

subplot(4,3,12)

```

```

for i=1:sumsum

```

```

    if A(i)==1

```

```

        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+r')

```

```

        hold on

```

```

    elseif A(i)==2

```

```

        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+g')

```

```

        hold on

```

```

    elseif A(i)==3

```

```

        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+b')

```

```

        hold on

```

```

    elseif A(i)==4

```

```

        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+k')

```

```

        hold on

```

```

    end
end

nnnn=num;
esi=1;
nc=3;
nnc=nc;
t=2;
r=10;
sumfasele(5,4)=0

cc(r,8,4)=0;my(5,5)=0;
[cc(1:nc,1:8,esi),u]=fcm(datatez(1:num(esi),1:8),nc);

for i=1:num(1)
    for xxx=1:nc
        faasele(xxx,i)=sqrt(sum((initialdata(i,:)-cc(xxx,esi)).^2));
    end

ppp(1,i)=u(1,i);
for ii=1:nc
    if u(ii,i)>=ppp(1,i)
        ppp(1,i)=u(ii,i);
        pp(1,i)=faasele(ii,i);
        khushe(1,i)=ii;
    end
end
sumfasele(1,khushe(1,i))=pp(1,i)+sumfasele(1,khushe(1,i));

end

tagh(1:nc,1:8,esi)=cc(1:nc,1:8,esi);taghnum(1)=0;

```

```

pp=0;
for i=1:num(esi)
    pp(1,i)=u(1,i);
    for ii=1:nc
        if u(ii,i)>=pp(1,i)
            pp(1,i)=u(ii,i);
            maxxx(1,i)=ii;
        end
    end
end
end

```

```

my(esi,1)=0;my(esi,2)=0;my(esi,3)=0;my(esi,4)=0;my(esi,5)=0;
for i=1:num(esi)
    if maxxx(1,i)==1
        my(esi,1)=my(esi,1)+1;
    elseif maxxx(1,i)==2
        my(esi,2)=my(esi,2)+1;
    elseif maxxx(1,i)==3
        my(esi,3)=my(esi,3)+1;
    elseif maxxx(1,i)==4
        my(esi,4)=my(esi,4)+1;
    elseif maxxx(1,i)==5
        my(esi,5)=my(esi,5)+1;
    end
end
end

```

```

subplot(4,3,1)
for i=1:numm
    if datatez(i,:)==0
        if maxxx(i)==1
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+r')
            hold on
        elseif maxxx(i)==2

```

```

        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+g')
        hold on
elseif maxxx(i)==3
        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+b')
        hold on
elseif maxxx(i)==4
        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+k')
        hold on
elseif maxxx(i)==5
        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'y')
        hold on
end
end

end

```

```

subplot(4,3,7)
for i=1:numm
    if datatez(i,:)~=0
        if maxxx(i)==1
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+r')
            hold on
        elseif maxxx(i)==2
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+g')
            hold on
        elseif maxxx(i)==3
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+b')
            hold on
        elseif maxxx(i)==4
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+k')
            hold on
        elseif maxxx(i)==5
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'y')

```

```

        hold on
    end
end

end

kkk=2;zzz=0;n1(8,4)=0;aaa=1;
nuumm=num(1);puyasumfasele(4,4)=0;
for esi=1:4

[ccc(1:nc,1:8,esi),uw]=fcm(datatez(1:nuumm,1:8),nc);

for i=1:nuumm+num(esi+1)
    for xxx=1:nc
        faasele(xxx,i)=sqrt(sum((initialdata(i,:)-ccc(xxx,esi)).^2));
    end
    if i>nuumm
        pp(1,i)=faasele(1,i);
    for ii=1:nc
        if faasele(ii,i)<=pp(1,i)
            pp(1,i)=faasele(ii,i);
            khushe(1,i)=ii;
        end
    end
end
    elseif i<=nuumm
        ppp(1,i)=uw(1,i);
    for ii=1:nc
        if uw(ii,i)>=ppp(1,i)
            ppp(1,i)=uw(ii,i);
            pp(1,i)=faasele(ii,i);
            khushe(1,i)=ii;
        end
    end
end
end

```

```

    end
sumfasele(esi+1,khushe(1,i))=pp(1,i)+sumfasele(esi+1,khushe(1,i));
if khushe(1,i)==1
    n1(aaa+1,1)=n1(aaa+1,1)+1;
elseif khushe(1,i)==2
    n1(aaa+1,2)=n1(aaa+1,2)+1;
    elseif khushe(1,i)==3
    n1(aaa+1,3)=n1(aaa+1,3)+1;
    elseif khushe(1,i)==4
    n1(aaa+1,4)=n1(aaa+1,4)+1;
end

if i<=nuumm
    if khushe(1,i)==1
        n1(aaa,1)=n1(aaa,1)+1;
    elseif khushe(1,i)==2
        n1(aaa,2)=n1(aaa,2)+1;
    elseif khushe(1,i)==3
        n1(aaa,3)=n1(aaa,3)+1;
    elseif khushe(1,i)==4
        n1(aaa,4)=n1(aaa,4)+1;
    end
end

end

end

subplot(4,3,esi+7)
for i=1:nuumm+num(esi+1)
    if datatez(i,:)=0
        if khushe(1,i)==1
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+r')
            hold on

```

```

elseif khushe(1,i)==2
    plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+g')
    hold on
elseif khushe(1,i)==3
    plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+b')
    hold on
elseif khushe(1,i)==4
    plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+k')
    hold on
elseif khushe(1,i)==5
    plot3(datatez(i,2),datatez(i,3),datatez(i,5),'y')
    hold on
end
end

end

nummm=nummm+nnnn(esi+1);aaa=aaa+2;

[cc(1:nc,1:8,esi+1),u1]=fcm(datatez(1:num(esi+1)+numm,1:8),nc);

tagh(1:nc,1:8,kkk)=cc(1:nc,1:8,esi+1);taghnum(kkk)=esi;kkk=kkk+1;
u=u1;

maxxue=max(u1);

pp=0;dd=0;
for i=1:numm+num(esi+1)
    pp(1,i)=u1(1,i);
    for ii=1:nc
        if u1(ii,i)>=pp(1,i)
            pp(1,i)=u1(ii,i);
            maxx(1,i)=ii;
        end
    end
end

```



```

        end
    end
end

z=0;
for i=1:nc
    for ii=1:nc
        if i~=ii
            z=z+1;
            oghcluster(z,1)=sqrt((cc(i,1,esi)-cc(ii,1,esi))^2+(cc(i,2,esi)-cc(ii,2,esi))^2+(cc(i,3,esi)-
cc(ii,3,esi))^2+(cc(i,4,esi)-cc(ii,4,esi))^2+(cc(i,5,esi)-cc(ii,5,esi))^2+(cc(i,6,esi)-
cc(ii,6,esi))^2+(cc(i,7,esi)-cc(ii,7,esi))^2+(cc(i,8,esi)-cc(ii,8,esi))^2);
        end
    end
end

sss=0;

    for i=numm+1:num(esi+1)+numm
        l(1,i)=0;
        if abs(maxxue(1,i)-1/nc)<=0.05
            iii=maxxx(1,i);
            oghdatcluster(1,i)=sqrt((datatez(i,1)-cc(iii,1,esi))^2+(datatez(i,2)-
cc(iii,2,esi))^2+(datatez(i,3)-cc(iii,3,esi))^2+(datatez(i,4)-cc(iii,4,esi))^2+(datatez(i,5)-
cc(iii,5,esi))^2+(datatez(i,6)-cc(iii,6,esi))^2+(datatez(i,7)-cc(iii,7,esi))^2+(datatez(i,8)-
cc(iii,8,esi))^2);
            if oghdatcluster(1,i) > min(oghcluster)
                l(1,i)=1;
            end
        end
    end
    if l(1,i)==1
        sss=sss+1;
    end
end

```

```

end
newm=0;
if sss/num(esi+1)>=0.05
%   u1(3,:)=1;
    newm=newm+1;

end
if esi<t
    if newm~=0
        nnc=nc;
        nc=nc+1;

        [cc(1:nc,1:8,esi+1),u]=fcm(datatez(1:num(esi+1)+numm,1:8),nc);
        tagh(1:nc,1:8,kkk)=cc(1:nc,1:8,esi+1);taghnum(kkk)=esi;kkk=kkk+1;
pp=0;
for i=1:numm+num(esi+1)
    pp(1,i)=u(1,i);
    for ii=1:nc
        if u(ii,i)>=pp(1,i)
            pp(1,i)=u(ii,i);
            maxxx(1,i)=ii;
        end
    end
end
end
end

end
elseif esi>=t
    if newm~=0
        nnc=nc;
        nc=nc+1;

        [cc(1:nc,1:8,esi+1),u]=fcm(datatez(1:num(esi+1)+numm,1:8),nc);
        tagh(1:nc,1:8,kkk)=cc(1:nc,1:8,esi+1);taghnum(kkk)=esi;kkk=kkk+1;

```

```

        pp=0;
    for i=1:numm+num(esi+1)
        pp(1,i)=u(1,i);
        for ii=1:nc
            if u(ii,i)>=pp(1,i)
                pp(1,i)=u(ii,i);
                maxxx(1,i)=ii;
            end
        end
    end
end

end

ncc=nc;jjj=0;
for mm=1:nc
    for mmm=1:nnc
        if abs(cc(mm,1:8,esi+1)-cc(mmm,1:8,esi))<=.3
            if ncc>2
                ncc=ncc-1; dd=1;
                for nn=1:numm+num(esi+1)
                    if maxxx(1,nn)==mm
                        for bb=nn+1:sumsum
                            datatez(bb-1,:)=datatez(bb,:);
                        end
                        sumsum=sumsum-1;
                    end
                    if nn<=numm
                        numm=numm-1;
                    else
                        num(esi+1)=num(esi+1)-1;
                    end
                    zzz=zzz+1;
                end
            end
        end
    end
end

```

```

        end
        jjj=1;
        break;

    end

end

end
end
if jjj==1
    break;
end
end
nc=nncc;
if dd==1
    cc(:,esi+1)=0;
    [cc(1:nncc,1:8,esi+1),u]=fcm(datatez(1:num(esi+1)+numm,1:8),nncc);
    tagh(1:nncc,1:8,kkk)=cc(1:nncc,1:8,esi+1);taghnum(kkk)=esi;kkk=kkk+1;
pp=0;
for i=1:numm+num(esi+1)
    pp(1,i)=u(1,i);
    for ii=1:nncc
        if u(ii,i)>=pp(1,i)
            pp(1,i)=u(ii,i);
            maxxx(1,i)=ii;
        end
    end
end
end
end

end

e=0;

```

```

subplot(4,3,esi+1)
for i=1:numm+num(esi+1)
    if datatez(i,:)~=0
        if maxxx(i)==1
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+r')
            hold on
        elseif maxxx(i)==2
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+g')
            hold on
        elseif maxxx(i)==3
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+b')
            hold on
        elseif maxxx(i)==4
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+k')
            hold on
        elseif maxxx(i)==5
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'y')
            hold on
        end
    end
end

```

end

```
numm=numm+num(esi+1);
```

```

for i=1:numm
    if maxxx(1,i)==1
        my(esi+1,1)=my(esi+1,1)+1;
    elseif maxxx(1,i)==2
        my(esi+1,2)=my(esi+1,2)+1;

    elseif maxxx(1,i)==3

```

```

    my(esi+1,3)=my(esi+1,3)+1;

elseif maxxx(1,i)==4
    my(esi+1,4)=my(esi+1,4)+1;
elseif maxxx(1,i)==5
    my(esi+1,5)=my(esi+1,5)+1;
end
end
for i=1:numm
    puyasumfasele(esi,maxxx(1,i))=pp(1,i)+puyasumfasele(esi,maxxx(1,i));
end

end

datatez=initialdata;
[cccc(1:3,1:8),u1]=fcm(datatez(1:3000,1:8),3);
for i=1:3000
    pp(1,i)=u1(1,i);
    for ii=1:3
        if u1(ii,i)>=pp(1,i)
            pp(1,i)=u1(ii,i);
            maxxx(1,i)=ii;
        end
    end
end
end
subplot(4,3,6)

for i=1:3000
    if datatez(i,:)~=0
        if maxxx(i)==1
            plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+r')
            hold on
        elseif maxxx(i)==2

```

```

        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+g')
        hold on
elseif maxxx(i)==3
        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+b')
        hold on
elseif maxxx(i)==4
        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'+k')
        hold on
elseif maxxx(i)==5
        plot3(datatez(i,2),datatez(i,3),datatez(i,5),'y')
        hold on
end
end

```

```

end
yy1=0;yy2=0;yy3=0;yy4=0;
for i=1:3000
    if maxxx(1,i)==1
        yy1=yy1+1;
    elseif maxxx(1,i)==2
        yy2=yy2+1;

    elseif maxxx(1,i)==3
        yy3=yy3+1;

    elseif maxxx(1,i)==4
        yy4=yy4+1;
    end
end
end

```

منابع فارسی

۱. ابطحی ترانه و اکبرزاده توتونچی محمدرضا، (۱۳۹۱)، "بهینه سازی الگوریتم خوشه بندی فازی با استفاده از جمعیت پویا و الگوهای مبتنی بر جمعیت"، یازدهمین کنفرانس سیستم های هوشمند ایران
۲. اسماعیل پور حسن و غفاری آشتیانی پیمان، (۱۳۸۱)، بازاریابی، انتشارات دانشگاه آزاد، اراک، ص ۸۵.
۳. خان بابایی محمد و زین العابدینی سیده فاطمه، (۱۳۹۲)، "مدل بکارگیری تکنیک های داده کاوی در شناسایی، بخش بندی و تحلیل رفتار مشتریان خدمات بانکداری الکترونیکی"، فصلنامه علمی - پژوهشی تحقیقات بازاریابی نوین، سال سوم، شماره ۲، صص ۱۷۵-۱۸۸.
۴. خیری بهرام و احمدی الهام، (۱۳۹۳)، "بررسی عوامل موثر بر قصد تغییر رفتار مشتریان و پیامدهای آن"، فصلنامه توسعه مدیریت پولی و بانکی، سال دوم، شماره ۲، صص ۱۰۱-۱۲۶.
۵. سپهری محمد مهدی و کارگری مهرداد، (۱۳۹۱)، "بهبود الگوریتم خوشه بندی مشتریان برای توزیع قطعات یدکی با رویکرد داده کاوی (k-means)"، نشریه بین المللی مهندسی صنایع و مدیریت تولید، شماره ۲، جلد ۲۳، صص ۲۴۹-۲۴۰.
۶. شوندی حسن، (۱۳۸۵)، "نظریه های مجموعه های فازی و کاربرد آن در مهندسی صنایع و مدیریت"، تهران: گسترش علوم پایه.
۷. طهماسبی پژمان، (۱۳۹۰)، "خوشه بندی داده های محیطی".

۸. غضنفری حسن، ملک محمدی سمیرا، علیزاده سمیه و فتح الله مهدی، (۱۳۸۹)، "بخش بندی مشتریان صادراتی میوه های خوراکی"، فصلنامه پژوهشنامه بازرگانی، شماره ۵۵، صص ۱۵۱-۱۸۱.
۹. قاسمی محمد سعید، خانگلدی مسعود، (۱۳۸۸)، "کاربرد منطق فازی در بازشناسی الگو: خوشه بندی"، دانشکده فنی و مهندسی، دانشگاه اراک
۱۰. مرتضوی سعید، آسمان دره یاسر، نجفی سیاهرودی مهدی و علوی سید مسلم، (۱۳۹۰)، "بخش بندی بازار گوشی تلفن همراه بر مبنای مزایای مورد انتظار مشتریان"، مدیریت بازرگانی، دوره سوم، شماره ۸، صص ۱۱۵-۱۳۲.
۱۱. میرواحدی سید سعید و نوراله اوغلی رقیه، (۱۳۹۰)، "تکنیک های تحلیل در تحقیقات بازاریابی"، انتشارات آیپژ.

1. Aggarwal, C. (2007), "**Data Streams: Models and Algorithms**", Advances in Database Systems, 31.
2. Alpaydin E, (2004), "**Introduction to Machine Learning**", the MIT Press.
3. Ann Maharaj, E, and D'Urso, P, (2011), "**Fuzzy clustering of time series in the frequency domain**", information sciences, 181:1187-1188 .
4. Berry M and Linoff G, (1997), "**Data Mining Techniques: For Marketing, Sales, and Customer Support**", New York: John Wiley and Sons.
5. Berson A, Smith S and Thearling K, (2004), "**Building Data Mining Applications for CRM**", Tata McGraw-Hill, New York.
6. Bezdek J.C., Keller J, Krishnapuram R., and Pal N. R., (1999), "**Fuzzy Models and Algorithms for Pattern Recognition and Image Processing**", Kluwer, Boston, London, Dordrecht.
7. Chuang H. M, Shen C.C., (2008), "**A study on the application of data mining techniques to enhance customer lifetime value-based on the department store industry**", The Seventh International Conference on Machine Learning and Cybernetics: 168-173.
8. Chunfei Zh, Zhiyi F, (2013), "**An Improved K-means Clustering Algorithm**", Journal of Information & Computational Science 10,1: 193–199

9. George J, Klir Bo Yuan,(1995)," **Fuzzy sets and Fuzzy Logic, Theory and applications**", Prentice Hall PTR.
10. Han J and Kamber M, (2001)," **Data Mining: Concepts and Techniques**", Morgan Kaufmann Publishers,San Francisco,USA.
11. Hosseininezhad F and Salajegheh A, (2012)," **Study and Comparison of Partitioning Clustering Algorithms**", Iranian Journal of Medical Informatics, 2, 1: 38-40.
12. Hesieh N.C , (2004)," **An integrated data mining and behavioral scoring model for analyzing bank customers**", 27: 623-633
13. Karypis G, Han E.-H, and Kumar V, (1999)," **Chameleon: ‘A hierarchical Clustering Algorithm Using Dynamic Modeling**” ,IEEE Computer, Special Issue on Data Analysis and Mining 32, 8: 68–75
14. Kasabov N.K, and Song Q, (2002), “**DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction**”, IEEE Transactions on Fuzzy Systems 10, 2: 144–154
15. Kim K-j, Ahn, H,(2008)," **A recommender system using GA K-means clustering in an online shopping market**", Expert Systems with Applications,34: 1200-1209.
16. Larose D.T,(2005)," **Discovering knowledge in Data , an introduction to Data mining** “.New Jersey . WILLEY.

17. Lee C-Y, Ma L, and Antonsson, E K, (2001),”**Evolutionary and adaptive synthesis methods**”, In Formal Engineering Design Synthesis (E. K. Antonsson and J. Cagan, eds), Cambridge University Press, Cambridge, U.K.: 270–320.
18. Lee Marcus T.H (2003),”**A Bayesian neural network model of consumer choice. Dissertation for the degree of Doctor of philosophy**”, University of Toronto.
19. Li C, Biswas G, Dale M, and Dale P, (2002), “**Matryoshka: a HMM based temporal data clustering methodology for modeling system dynamics**”, Intelligent Data Analysis, 6, 3: 281–308
20. Li Der-Chiang, Dai Wen-Li and Tseng Ean-Ting, (2011),”**A twostage clustering mehod to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business**”, expert systems with applications, 14: 1-6.
21. Li W, Wu X, (2010),”**Credit Card Customer Segmentation and Target Marketing Based on Data Mining**”, International Conference on Computational Intelligence and Security: 73 – 76.
22. Liao S.H., Chen Y.J., Hsieh H.H., (2011), “**mining customer knowledge for direct selling and marketing, Expert Systems with Applications**”, 38: 6059–6069.
23. Linder R, Geier J, Kolliker M, (2004),” **Artificial neural networks, classification trees and regression: which method for which customer**

- base, Journal of Database Marketing & Customer Strategy Management**", 11, 4: 344-356.
24. Malthouse E, C, (2001), "**Assessing the performance of direct marketing scoring models**", Journal of Interactive Marketing, 15: 49-62.
 25. Mitchell, Tom M, (1997), "**Machine Learning**", McGraw-Hill Science.
 26. Nanda S.R., Mahanty B., Tiwari M.K,(2010), "**Clustering Indian stock market data for portfolio management**", Expert Systems with Applications, 37,8793–8798.
 27. Peters G, Weber R, Nowatzke R, (2012), "**Dynamic rough clustering and its applications**", Applied Soft Computing, 12: 3193-3195.
 28. Robert S, Gyrodi,"**A Comparative Study of Iterative Algorithm in Association Rules Mining** ", Studies in Information and control, 12, 3.
 29. Sander J,(2003), "**Principles of Knowledge Discovery in Data: Clustering I**", Department of Computing Science University of Alberta, Tutorial Slides.
 30. Seret A, Seppe K.L.M. Broucke Y, Baesens B, Vanthienen J,(2014), "**A dynamic understanding of customer behavior processes based on clustering and sequence mining**", Expert Systems with Applications,41: 4648-4657.
 31. Weber R,(2007), Fuzzy Clustering In Dynamic Data Mining- Techniques And Applications,pp:315-330,In "**Advances in fuzzy clustering and its applications**", Valente de Oliveira J, Pedrycz W, Wiley.

32. Wu Hsin-Hung, Chang En-Chi and Lo Chiao-Fang, (2009),” **Applying RFM model and K-means method in customer value analysis of an outfitter**, International Conference on Concurrent Engineering, New York.
33. You C, (2009),”**On the convergence of uncertain sequences, Mathematical and Computer Modeling**”, 49: 482-487.
34. Zhenguo Z, Ying K, Gen P, (2012),”**Dynamic data mining based on cloud model**”, Atlantis press.
35. <http://ceit.aut.ac.ir/~shiry>, Dr.Saeed Shiry Ghidary Academic website
36. <http://www.thearling.com/> an introduction to data mining.htm.

Abstract

Today, customer is considered as the main key of success or unsuccess a company. Therefore the study of the customers' behavior has been regarded by marketing researchers during recent decades. The analysis of the customers' loyalty, satisfaction and preferences are the most important issues in this area that researchers try to analysis of them by using different tools and methods. This study in terms of objective is applied research, in terms of methodology is survey research and in terms of subject, is marketing research. The purpose of this research is the customers' segmentation using data mining techniques and the prediction of their behavior using dynamic C-means algorithm. For this purpose, the purchase data of 3000 Samsung mobile phone users has received and then this data has considered by Matlab 2008a after writing C-means algorithm. The results show that dynamic technique presents more real segmentations than static technique. Moreover the analysis of behaviors during a time period shows that customers are tendered to the price of the mobile phones. Traits such as the quality of the camera and processor and the number of sim cards and RAM are the reasons of change in the customers' segmentations.

Key words: Customer behavior, Dynamic data mining, Clustering, FCM



University of Shahrood

Faculty of Industries Engineering & Management

**Fuzzy clustering of customers and analysis of their
behavior using dynamic data mining approach**

(Case Study: Samsung Mobile Phones)

Maryam Alaeizadeh

Supervisor:

Dr. Reza Sheikh

February 2015