

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

سمینار

عنوان
روش‌های خوشه بندی فازی

نگارنده:

۵	چکیده
۶	فصل اول: مقدمه
۷	۱-۱- مقدمه
۱۰	فصل دوم: مطالعه خوشه بندی فازی
۱۱	۱-۲- مقدمه
۱۲	۲-۲- هدف از خوشه بندی
۱۳	۳-۲- تفاوت خوشه بندی با طبقه بندی
۱۴	۴-۲- نظریه مجموعه های فازی
۱۴	۵-۲- مفهوم خوشه بندی فازی
۱۵	۲-۵-۱- تفاوت خوشه بندی فازی با خوشه بندی کلاسیک
۱۷	۲-۱۱- جمع بندی
۱۸	فصل سوم: بررسی و ارزیابی روش های خوشه بندی فازی
۱۹	۳-۱- مقدمه
۲۰	۳-۲- روش خوشه بندی فازی FCM
۲۱	۳-۳- الگوریتم های پایه ای خوشه بندی فازی
۲۱	۳-۴- الگوریتم خوشه بندی c میانگین
۲۶	۳-۴-۱- الگوریتم c-mean فازی جریمه شده
۲۷	۳-۴-۲- الگوریتم c-mean فازی متناوب
۲۸	۳-۴-۳- الگوریتم possibilistic c-mean فازی
۳۰	۳-۴-۴- الگوریتم (KFCM) Kernel fuzzy c-mean
۳۱	۳-۴-۵- الگوریتم خوشه بندی c میانگین برای داده های نویزی
۳۲	۳-۴-۶- الگوریتم خوشه بندی c میانگین با استفاده از نمونه های برچسب گذاری شده
۳۲	۳-۴-۷- الگوریتم خوشه بندی c میانگین مبتنی بر آنتروپی
۳۳	۳-۴-۸- الگوریتم خوشه بندی c میانگین مبتنی بر آنتروپی برای داده های نویزی

۳-۴-۹	الگوریتم خوشه بندی C میانگین با استفاده از یادگیری وزن ویژگی ها	۳۴
۳-۶-۶	الگوریتم های حاصل از تغییر در تابع فاصله	۳۵
۳-۶-۱	Gustafson-Kessel الگوریتم	۳۵
۳-۶-۲	Fuzzy shell clustering الگوریتم	۳۶
۳-۶-۳	الگوریتم های خوشه بندی فازی مبتنی بر هسته	۳۷
۳-۷-۷	الگوریتم های حاصل از تغییر در تابع هدف	۳۷
۳-۷-۱	Noise Handling Variants روش	۳۸
۳-۷-۲	Fuzzifier Variants روش	۳۸
۳-۷-۳	Cluster Number Determination Variants روش	۳۸
۳-۷-۴	Possibilistic C-Means صورت های دیگر	۳۸
۳-۷-۴-۱	دفاعه خوشه‌ای	۳۸
۳-۷-۴-۲	صور تغییر یافته PCM، مبتنی بر ترکیب FCM و PCM	۳۹
۳-۸-۸	روش خوشه بندی مستقل از مقیاس بر پایه FCM	۳۹
۳-۹-۹	روش های خوشه بندی وفقی	۴۰
۳-۹-۱	الگوریتم های خوشه بندی فازی بر پایه یک فاصله وفقی واحد	۴۰
۳-۹-۱-۱	روش خوشه بندی فازی بر پایه یک فاصله مجذور وفقی تعریف شده	۴۰
۳-۹-۱-۲	روش خوشه بندی فازی بر پایه یک فاصله مجذور وفقی	۴۱
۳-۹-۲	الگوریتم های خوشه بندی فازی بر پایه فواصل وفقی متفاوت برای خوشه های مختلف	۴۱
۳-۹-۲-۱	روش خوشه بندی فازی گوستاوسن-کسل (GK)	۴۲
۳-۹-۲-۲	روش خوشه بندی فازی بر پایه یک فاصله مجذور وفقی	۴۲
۳-۱۰-۱۰	معیارهای کارایی	۴۲
۳-۱۱-۱۱	جمع بندی	۴۸

فصل چهارم: نتیجه گیری و پیشنهادات ۴۹

۴-۱-۱	نتیجه گیری و پیشنهادات	۵۰
-------	------------------------	----

مراجع		۵۱
-------	--	----

فهرست اشکال

- شکل ۱-۲- خوشه‌بندی نمونه‌های ورودی ۱۱
- شکل ۲-۲- خوشه بندی وسایل نقلیه ۱۲
- شکل ۳-۲- مجموعه داده پروانه‌ای ۱۶
- شکل ۴-۲- خوشه بندی فازی داده ۱۶
- شکل ۵-۳- توزیع یک بعدی نمونه ها ۲۳
- شکل ۶-۳- خوشه بندی کلاسیک نمونه های ورودی ۲۴
- شکل ۷-۳- خوشه بندی فازی نمونه ها ۲۴
- شکل ۸-۳- معیارهای تشابه بر اساس توابع فاصله مختلف ۲۵
- شکل ۹-۲- ایراد وارد بر ۳ درجه عضویت احتمالی ۲۸

چکیده

در الگوریتم خوشه‌بندی فازی FCM باید تعداد و مراکز خوشه‌ها توسط کاربر در ابتدا مشخص شوند. کیفیت این الگوریتم بشدت به تعداد اولیه خوشه‌ها و مکان اولیه مراکز خوشه‌ها بستگی دارد. هدف از خوشه‌بندی فازی استخراج مدل‌های فازی از داده‌هاست. کاربردهای متعدد خوشه‌بندی فازی در تحلیل داده‌ها و تشخیص الگو و نیز زمینه‌های پژوهشی موجود در این زمینه از جمله استفاده از آن در حل مسائل مسیریابی، تخصیص و زمان‌بندی نیاز به مطالعه الگوریتم‌های موجود و بهبود و اصلاح آن‌ها را آشکارتر می‌نماید. در این سمینار چندین روش خوشه‌بندی فازی بر پایه روش خوشه‌بندی استاندارد FCM مورد بررسی و مطالعه قرار گرفته است.

کلمات کلیدی: خوشه بندی، خوشه بندی فازی، C میانگین، طبقه‌بندی

فصل اول

مقدمه

خوشه‌بندی یک کلاس‌بندی بدون نظارت است که در آن کلاس‌ها از پیش تعریف نشده‌اند. هدف از انجام خوشه‌بندی ارائه چشم انداز مناسبی از اتفاقات در حال وقوع در پایگاه داده‌ها به مصرف کننده نهایی اطلاعات می‌باشد. کاربرد دیگر خوشه‌بندی را می‌توان در تعیین داده‌هایی که با سایر داده‌ها تفاوت چشمگیر دارند عنوان نمود. در خوشه‌بندی کلاسیک، هر نمونه ورودی متعلق به یک و فقط یک خوشه است و نمی‌تواند عضو دو خوشه و یا بیشتر باشد. به عبارتی خوشه‌ها همپوشانی ندارند در حالیکه در خوشه‌بندی فازی یک نمونه می‌تواند متعلق به بیش از یک خوشه باشد. خوشه‌بندی فازی به کشف مدل‌های فازی از داده‌ها می‌پردازد. این مقاله به بررسی الگوریتم‌هایی که هدفشان استخراج خوشه‌های فازی از داده‌های قطعی است، می‌پردازد. یکی از اولین روش‌های خوشه‌بندی فازی که بر مبنای تابع هدف و استفاده از فاصله اقلیدسی بنا شده بود در سال ۱۹۷۴ توسط دان ۳ ارائه و سپس توسط بزدک ۴ تعمیم داده شد. پس از آن یانگ ۵ یک بررسی اجمالی روی روش‌های خوشه‌بندی فازی انجام داد. سپس گوستافسون و کسل ۶ در سال ۱۹۷۹ الگوریتم خوشه‌بندی فازی با استفاده از ماتریس کوواریانس فازی را ارائه نمودند. تلفیق رویکرد امکان در خوشه‌بندی فازی نخستین بار توسط کیم و کریشناپورام ۷ در سال ۱۹۹۳ صورت گرفت. پس از آن نیز اصلاحات و بهبودهای بسیاری روی الگوریتم‌های ارائه شده صورت گرفته‌اند که در این تحقیق به برخی از آن‌ها خواهیم پرداخت. تحلیل خوشه‌بندی به دنبال سازمان دهی مجموعه‌ای از داده‌ها در یک سری خوشه است به طوری که داده‌ها (که معمولاً هر کدام نشان دهنده برداری از مقادیر کمی در یک فضای چندین بعدی است)، در هر خوشه بالاترین درجه شباهت را دارا بوده و داده‌های متعلق به خوشه‌های مختلف دارای حداکثر درجه عدم شباهت هستند. روش‌های تحلیل خوشه‌ها به دو روش عمده سلسله مراتبی و تفکیکی تقسیم می‌شوند. روش‌های سلسله‌مراتبی، کاملاً سلسله‌مراتبی عمل می‌کنند به این معنی که داده‌های ورودی در یک دنباله مرتب قرار می‌گیرند. روش‌های سلسله‌مراتبی می‌توانند تراکمی یا تفرقی باشند. در روش تراکمی دنباله‌ای از بخش‌ها ایجاد می‌شود که از خوشه‌بندی که در آن هر داده در یک خوشه قرار می‌گیرد شروع و با خوشه‌یابی که در آن تمام داده‌ها در یک خوشه قرار می‌گیرند، خاتمه می‌یابد. در روش تفرقی ابتدا همه داده‌ها در یک خوشه قرار می‌گیرند و سپس عملیات شکستن تا رسیدن به یک حد توقف که معمولاً پیش از رسیدن به خوشه‌هایی با یک عضو می‌باشد، ادامه می‌یابد. تقسیم‌بندی یک مجموعه داده به تعداد خوشه از پیش تعیین شده مبحث بسیار مهمی در تحلیل داده، بازشناسی الگو و پردازش تصویر محسوب می‌شود. در روش‌های تفکیکی، به دست آوردن یک تفکیک از داده‌های ورودی در تعداد مشخصی از خوشه‌ها مورد نظر است. چنین روش‌هایی معمولاً دنبال تفکیکی هستند که یک تابع کارایی را (معمولاً به صورت محلی) بهینه کند. برای بهبود کیفیت خوشه‌بندی الگوریتم چندین بار و در

نقاط شروع مختلف اجرا می شود و بهترین وضعیت بدست آمده از کل دفعات اجرا به عنوان خروجی خوشه بندی انتخاب می شود. روش های تفکیکی به دو بخش خوشه بندی سخت و خوشه بندی فازی تقسیم می شوند. در خوشه بندی سخت، هر داده به یک و فقط یک خوشه نسبت داده می شود، در حالیکه در خوشه بندی فازی، یک تفکیک فازی صورت می گیرد به این معنی که هر داده با یک درجه تعلق به هر خوشه متعلق است. در شرایط واقعی خوشه بندی فازی بسیار طبیعی تر از خوشه بندی سخت است، چون داده های موجود در مرز خوشه های مختلف مجبور به تعلق کامل به یکی از خوشه ها نیستند، و با درجه تعلق بین 0 تا 1 که نشان دهنده تعلق نسبی آنها است تفکیک می شوند. الگوریتم های خوشه یابی دینامیک، الگوریتم های تکراری دو مرحل های جایگزین الگوریتم های خوشه بندی سخت هستند که شامل ساخت خوشه ها در هر تکرار و شناسایی نماینده های مناسب یا مراکز خوشه ها برای هر خوشه با بهینه سازی محلی یک معیار کارایی بین خوشه ها و نماینده های مربوط به آنها می باشند. این پروسه بهینه سازی از مجموعه ای از نماینده های اولیه یا یک تقسیم بندی اولیه شروع شده و به صورت متقابل یک مرحله تخصیص (که در آن نماینده ها ثابتند) را در جهت اختصاص دادن الگوها به خوشه ها بر طبق شباهت آنها به نماینده ها به کار می برد. سپس در مرحله نمایش (که در آن تفکیک ثابت است) نماینده ها بر طبق تخصیص الگوها در مرحله قبل به روز می شوند و این مراحل تا زمان همگرایی الگوریتم که در آن معیار کارایی به یک مقدار ایستا برسد ادامه می یابند. الگوریتم خوشه یابی سخت دینامیک وفقی نیز یک معیار را بر پایه یک میزان تناسب بین خوشه ها و نماینده های آنها بهینه می کند با این تفاوت که فاصله های مورد استفاده برای مقایسه خوشه ها و نماینده های آنها در هر تکرار تغییر می کنند. این فاصله ها به طور قطعی معین نمی شوند و می توانند برای خوشه های مختلف متفاوت باشند. فایده این فاصله های وفقی این است که الگوریتم خوشه یابی قادر به تشخیص خوشه ها با اشکال و اندازه های مختلف است. مقداردهی اولیه، مرحله تخصیص، و حد توقف تقریباً در الگوریتم های خوشه یابی سخت دینامیک وفقی و غیر وفقی یکسان است. تفاوت عمده بین این الگوریتم ها در مرحله نمایش است که در حالت وفقی شامل دو قدم (به روز رسانی نماینده های خوشه ها و فاصله ها) است.

تئوری مجموعه فازی به کار گرفته شده در تحلیل خوشه یابی عمدتاً بر روی خوشه یابی فازی بر پایه روابط فازی و توابع هدف تمرکز دارد. یکی از اولین روش های خوشه یابی فازی، بر پایه یک تابع هدف، و با استفاده از فاصله اقلیدسی توسط dunn ارائه شده و بعدها توسط Bezdek تعمیم داده شد. Yang یک بررسی اجمالی قابل فهم بر روی روش های خوشه یابی فازی انجام داد. Jajuga و Greoenen یک مدل خوشه یابی فازی بر پایه ریشه توان دوم فاصله مینکوفسکی شامل فاصله اقلیدسی مجذور شده و مجذور نشده معرفی کردند. اولین الگوریتم خوشه یابی فازی وفقی بر پایه فواصل درجه دوم تعریف شده با یک ماتریس کوواریانس فازی، توسط Gustafson و Kessel بیان شد. مطالعه با جزئیات بیشتر روی این

الگوریتم اولیه توسط Krishnapuram و Kim انجام شد. Babuska و همکارانش تخمین ماتریس کوواریانس فازی مورد استفاده در الگوریتم (GK) Gustafson-Kessel را بهبود بخشیدند. پس از آن Frigni و Nasraoui روشی بر پایه الگوریتم استاندارد FCM برای حل مشکل خوشه بندی و وزن دهی ویژگی ها به طور همزمان ارائه کردند. در این مقاله روش های مختلف خوشه بندی فازی تفکیکی بر پایه فواصل مجذور و فقی ارائه و مقایسه شده اند. این فواصل مجذور و فقی توسط ماتریس های معین مثبت مقارنی تعریف می شوند که برای به روزسازی فواصل در قدم دوم مرحله نمایش الگوریتم، باید معکوس شوند. گرچه در حالاتی که این ماتریس ها مجازاً تکین باشند، مشکلات عددی ممکن است رخ دهند و ماتریس معکوس پذیر نباشد. یک روش برای حل این مشکل استفاده از روش ارائه شده توسط Babuska و همکارانش است.