



دانشگاه زنجان

پایان نامه کارشناسی

دانشگاه سراسری زنجان

رشته : مهندسی برق

گرایش : کنترل

عنوان

خوشه بندی راداری (خوشه بندی چند بعدی)

استاد راهنما : جناب آقای دکتر رضا امیدی

نگارش : مجید محمدی

90611250

بهمن ماه 1396

فهرست مطالب

عنوان	صفحه
۱	۵
بررسی الگوریتم‌های خوشه‌بندی چندبعدي	۵
۱-۱	۶
مقدمه	۶
۱-۱-۱	۷
خوشه‌بندی چیست؟	۷
۱-۱-۲	۸
خوشه‌بندی در مقابل طبقه‌بندی	۸
۱-۱-۳	۹
یادگیری با نظارت در مقابل یادگیری بدون نظارت	۹
۱-۱-۴	۱۰
کاربردها:	۱۰
۱-۱-۵	۱۱
هدف از خوشه‌بندی چیست؟	۱۱
۱-۱-۶	۱۱
خوشه‌بندی در مقابل چندی‌سازی برداری:	۱۱
۲-۱	۱۳
تعاریف اولیه	۱۳
۱-۲-۱	۱۳
ماتریس داده‌ها	۱۳
۲-۲-۱	۱۳
روشهای نرمالیزاسیون بردارهای ویژگی	۱۳
۳-۲-۱	۱۴
ماتریس مجاورت	۱۴
۴-۲-۱	۱۵
روشهای اندازه‌گیری مجاورت دو بردار	۱۵
۳-۱	۱۸
الگوریتم‌های خوشه‌بندی تقسیمی	۱۸
۱-۳-۱	۱۸
الگوریتم خوشه‌بندی K-means	۱۸
۲-۳-۱	۲۳
الگوریتم خوشه‌بندی K-medoid	۲۳
۳-۳-۱	۲۶
الگوریتم خوشه‌بندی فازی Fuzzy c-means	۲۶
۴-۳-۱	۳۲
الگوریتم ISODATA	۳۲
۴-۱	۳۷
الگوریتم‌های خوشه‌بندی سلسله‌مراتبی	۳۷
۱-۴-۱	۳۸
الگوریتم‌های پیوندی	۳۸
۲-۴-۱	۴۰
الگوریتم‌های تفکیکی	۴۰
۵-۱	۴۱
الگوریتم‌های خوشه‌بندی مبتنی بر توزیع چگالی	۴۱
۱-۵-۱	۴۴
الگوریتم DBSCAN	۴۴

فهرست شکلها

عنوان صفحه

شکل ۱-۱: خوشه بندی نمونه های ورودی..... ۷

شکل ۲-۱: خوشه بندی وسایل نقلیه..... ۸

شکل ۳-۱ (a): در طبقه بندی با استفاده یک سری اطلاعات اولیه داده ها به دسته های معلومی نسبت داده

می شوند. (b) در خوشه بندی داده ها با توجه به الگوریتم انتخاب شده به خوشه هایی نسبت داده

می شوند..... ۹

شکل ۴-۱۱: روند تغییر مراکز خوشه ها در الگوریتم K-means..... ۱۹

شکل ۵-۱: فلوجارت الگوریتم K-means..... ۲۱

شکل ۶-۱: مجموعه داده پروانه ای..... ۲۸

شکل ۷-۱: توزیع یک بعدی نمونه ها..... ۳۰

شکل ۸-۱: خوشه بندی کلاسیک نمونه های ورودی..... ۳۰

شکل ۹-۱: خوشه بندی فازی نمونه ها..... ۳۱

شکل ۱۰-۱: فلوجارت الگوریتم ISODATA..... ۳۶

شکل ۱۱-۱: دیاگرام درختی برای نمایش الگوریتم های سلسله مراتبی..... ۳۸

شکل ۱۲-۱: انواع الگوریتم های پیوندی بر اساس معیار اندازه گیری مشابهت یا عدم تشابه..... ۳۹

شکل ۱۳-۱: روش center-based برای تعریف چگالی هر نقطه..... ۴۲

شکل ۱۴-۱: انواع نقاط پایگاه داده در الگوریتم های مبتنی بر توزیع چگالی..... ۴۳

شکل ۱۵-۱: پایگاه داده های مورد استفاده برای تست الگوریتم های مبتنی بر توزیع چگالی..... ۴۳

شکل ۱۶-۱: خاصیت عدم تقارن قابلیت دسترسی مستقیم چگالی..... ۴۴

شکل ۱۷-۱: خاصیت عدم تقارن قابلیت دسترسی چگالی..... ۴۵

شکل ۱۸-۱: دو نقطه متصل چگالی..... ۴۵

شکل ۱۹-۱: خوشه بندی انجام شده با استفاده از الگوریتم DBSCAN..... ۴۷

شکل ۲۰-۱: مقدار $4-dist$ محاسبه شده بر روی یک پایگاه داده..... ۴۸

۱ بررسی الگوریتم‌های خوشه‌بندی چندبعدي

۱-۱ مقدمه

به طور کلی مسائل کلاسه‌بندی^۱ به دو دسته تقسیم می‌شوند. دسته اول مسائل تخصیص^۲

هستند که به روش یادگیری با مربی^۳ انجام می‌شوند. دسته دوم مسائل خوشه‌بندی^۴ هستند که بر پایه یادگیری بدون مربی^۵، نمونه‌ها را در دسته‌های محدودی تقسیم‌بندی می‌کنند به طوریکه نمونه‌های

موجود در هر خوشه کمترین پراکندگی را نسبت به قرار گرفتن در سایر خوشه‌ها دارند و اعضای هر

خوشه بیشترین شباهت را به هم دارند. در خوشه‌بندی از ابزارهای ریاضی برای تشخیص تشابه بین

اشیاء داخل یک کلاستر استفاده می‌شود که این ابزار می‌تواند در دو حالت کلی قطعی یا فازی باشند.

خوشه‌بندی قطعی یا سخت^۶ توسط یک تابع عضویت قطعی توصیف می‌شود. این تابع هر

شیء را به یک و فقط یکی از خوشه‌ها با درجه عضویت یک نسبت می‌دهد و لذا مرز بین خوشه‌ها

نیز کاملاً مجزا شده‌اند. البته الگوهای موجود در طبیعت اکثراً با هم تداخل دارند و نمی‌توان هر نمونه

را به صورت کاملاً مجزا متعلق به فقط یکی از مجموعه‌ها یا عوامل طبیعت دانست، لذا به نظر

می‌رسد که استفاده از خوشه‌بندی قطعی نمی‌تواند همواره و به خصوص در پدیده‌های طبیعی، میزان

تعلق دقیق الگوها به مجموعه‌ها را به طور واقعی نشان دهد.

خوشه‌بندی فازی بر مبنای نظریه فازی شکل گرفته است و بنا به خواص تئوری فازی، بین

کلاسترها مرزبندی دقیقی وجود ندارد و تابع عضویت هر الگو با درجه‌ای بین صفر و یک نشان داده

می‌شود. درجه عضویت هر نمونه نسبت به هر خوشه نشان می‌دهد که نمونه تا چه حد به این خوشه

تعلق دارد. بنابراین در خوشه‌بندی فازی، هر یک از الگوها متعلق به چندین خوشه با درجات عضویت

متفاوت هستند.

^۱ Classification

^۲ Assignment

^۳ supervised learning

^۴ Clustering

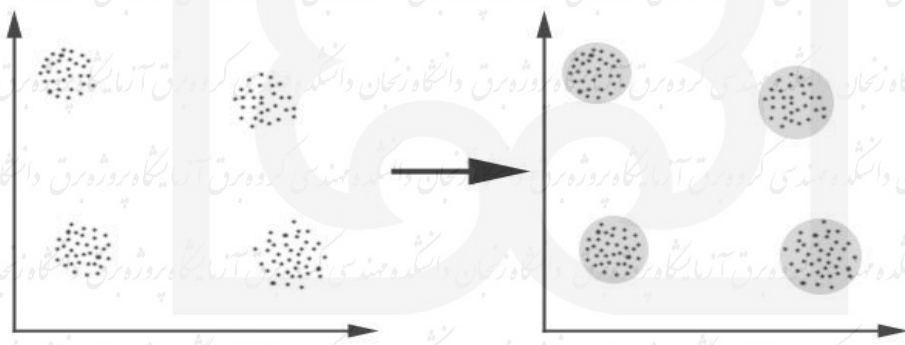
^۵ unsupervised learning

^۶ Crisp or Hard Clustering

۱-۱-۱ خوشه‌بندی چیست؟

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 اعضای مجموعه با یکدیگر مشابه بوده و با اشیاء موجود در خوشه‌های (مجموعه‌های) دیگر غیر مشابه می‌باشند. برای مشابه بودن می‌توان معیارهای مختلفی را در نظر گرفت مثلاً می‌توان معیار فاصله را برای خوشه‌بندی مورد استفاده قرار داد و اشیائی را که به یکدیگر نزدیکتر هستند را بعنوان یک خوشه در نظر گرفت که به این نوع خوشه‌بندی، خوشه‌بندی مبتنی بر فاصله نیز گفته می‌شود.

بعنوان مثالی از خوشه‌بندی، در شکل ۱-۱ نمونه‌های ورودی در سمت چپ به چهار خوشه‌بندی مشابه شکل سمت راست تقسیم می‌شوند. در این مثال هر یک از نمونه‌های ورودی به یکی از خوشه‌ها تعلق دارد و نمونه‌ای وجود ندارد که متعلق به بیش از یک خوشه باشد.



شکل ۱-۱: خوشه‌بندی نمونه‌های ورودی

بعنوان یک مثال دیگر شکل ۱-۲ را در نظر بگیرید در این شکل هر یک از دایره‌های کوچک یک وسیله نقلیه (شیء) را نشان می‌دهد که با ویژگی‌های وزن و حداکثر سرعت مشخص شده‌اند. هر یک از بیضی‌ها یک خوشه می‌باشد و عبارت کنار هر بیضی برچسب آن خوشه را نشان می‌دهد. کل

فهرست منابع ۱-۷

[۱] P. Tan, M. Steinbach and V. Kumar, *Introduction to Data Mining*, ۲۰۰۶.

[۲] S. Theodoridis and K. Koutroubass, *Pattern Recognition*, Elsevier, ۲۰۰۶.

[۳] M. Ester, H. Kriegel, J. Sander and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining, KDD*, ۱۹۹۶.

[۴] P. C. Wang and C. R. Ward, "Method for radar pattern recognition by sorting signals into data clusters," US Patent # ۷۰۳۴۷۳۸ B۱, ۲۰۰۶.

[۵] P. Hore, L. O. Hall and D. B. Goldgof, "Single Pass Fuzzy c-means," in *Proceedings of IEEE International Conference on Fuzzy Systems*, ۲۰۰۷.

[۶] S. Nascimento, B. Mirkin and F. Moura-Pires, "A Fuzzy Clustering Model of Data and Fuzzy c-Means," in *Proceedings of IEEE International Conference on Fuzzy Systems*, ۲۰۰۰.

[۷] A. K. Jain, M. N. Murty and P. J. Flynn, "Data Clustering: A review," *ACM Computing Surveys*, Vol. ۳۱, No. ۳, September ۱۹۹۹.

[۸] G. Fung, *A Comprehensive Overview of Basic Clustering Algorithms*, ۲۰۰۱.

[۹] A. Baraldi and P. Blonda, "A Survey of Fuzzy Clustering Algorithms for Pattern Recognition-Part I," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. ۲۹, No.۶, December ۱۹۹۹.

[۱۰] A. Baraldi and P. Blonda, "A Survey of Fuzzy Clustering Algorithms for Pattern Recognition-Part II," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. ۲۹, No.۶, December ۱۹۹۹.

فهرست مطالب

عنوان صفحه

1 خوشه بندی پالس های رادار 5

1-1 مقدمه 6

2-1 الگوریتم خوشه بندی ISODATA 7

3-1 الگوریتم خوشه بندی DBSCAN 13

4-1 دی اینترلیو کردن و پیوند خوشه های تشخیص داده شده 20

1-4-1 تشخیص PRI با استفاده از هیستوگرام DTOA 21

5-1 الگوریتم جامع در نظر گرفته شده جهت خوشه بندی پالس های رادار 25

6-1 نتایج شبیه سازی های انجام شده 29

7-1 فهرست منابع 35

فهرست شکلها

عنوان صفحه

شکل 1-1: فلوجارت الگوریتم ISODATA 12

شکل 1-2: خاصیت عدم تقارن قابلیت دسترسی مستقیم چگالی 14

شکل 1-3: خاصیت عدم تقارن قابلیت دسترسی چگالی 15

شکل 1-4: دو نقطه متصل چگالی 15

شکل 1-5: خوشه بندی انجام شده با استفاده از الگوریتم DBSCAN 17

شکل 1-6: مقدار $dist-4$ محاسبه شده بر روی یک پایگاه داده 18

شکل 1-7: محاسبه DTOA مرتبه 1، 2 و 3 22

شکل 1-8: رشته پالس تولید شده توسط رادار دارای Stagger-PRI مرتبه 3 23

شکل 1-9: الگوریتم جامع در نظر گرفته شده جهت خوشه بندی پالس های رادار 26

شکل 1-10: پایگاه داده مورد تست در نیمه اول بازه زمانی 29

شکل 1-11: پایگاه داده مورد تست در نیمه دوم بازه زمانی 30

شکل 1-12: نتیجه خوشه بندی الگوریتم DBSCAN پس از اولین بار پرشدن بافر 31

شکل 1-13: نتیجه خوشه بندی پس از تشخیص PRI و پیوند خوشه ها پس از اولین بار پرشدن بافر 32

شکل 1-14: نتیجه خوشه بندی الگوریتم جامع پس از دومین بار پرشدن بافر 33

شکل 1-15: نتیجه خوشه بندی الگوریتم جامع پس از سومین بار پرشدن بافر 34

1 خوشه بندی پالس های رادار

1-1 مقدمه

در این گزارش یک الگوریتم جامع جهت خوشه‌بندی پالس‌های رادار آشکارسازی شده مورد بررسی

قرار می‌گیرد. این الگوریتم جامع مبتنی بر الگوریتم‌های خوشه‌بندی $ISODATA^1$ یا $DBSCAN^2$ با

استفاده از $PDWs^3$ پالس‌های رادار و دی‌ایتریلیو کردن و ادغام خوشه‌ها با استفاده از PRI^4 پالس‌های

رادار می‌باشد. در الگوریتم جامع قابلیت خوشه‌بندی یک پالس دریافتی به محض ورود آن

(Tracking mode) نیز در نظر گرفته شده است، به طوری‌که همیشه نیاز به داشتن تعداد قابل

ملاحظه‌ای $PDWs$ جهت خوشه‌بندی نمی‌باشد. همچنین در این الگوریتم، در شرایطی که در

خروجی الگوریتم خوشه‌بندی $ISODATA$ یا $DBSCAN$ چند خوشه به عنوان یک خوشه تشخیص

داده شده باشند و یا یک خوشه به چند خوشه قسمت تقسیم شده باشد، با استفاده از روش‌های

آشکارسازی PRI ، خوشه‌های به هم پیوند خورده، تقسیم شده (دی‌ایتریلیو شده) و یا یک خوشه‌ای

که به چند قسمت تقسیم شده است، به هم پیوند می‌خورد. در این گزارش در ابتدا الگوریتم‌های

خوشه‌بندی $ISODATA$ و $DBSCAN$ توضیح داده می‌شود، سپس روش‌های تشخیص PRI بر

اساس هیستوگرام بدست آمده از $DTOA^5$ پالس‌های رادار معرفی می‌شود و در نهایت الگوریتم جامع

در نظر گرفته شده ارائه می‌شود.

¹ Iterative Self-Organizing Data Analysis Techniques

² Density-Based Spatial Clustering of Applications with Noise

³ Pulse Descriptive Words

⁴ Pulse Repetition Interval

⁵ Differential Time of Arrival

1-2 الگوریتم خوشه بندی ISODATA

الگوریتم ISODATA بر اساس الگوریتم K-means طراحی شده است با این تفاوت که با انجام یک

سری پردازش اضافی قابلیت تفکیک یک خوشه به دو خوشه و همچنین پیوند دو خوشه به یکدیگر

را در طی الگوریتم فراهم می‌آورد [2]. در الگوریتم K-means همانطور که اشاره شد، تعداد

خوشه‌ها، K ، به عنوان ورودی به الگوریتم داده می‌شود و در نهایت پایگاه داده به K خوشه متمایز

تقسیم می‌شود، درحالی‌که در بسیاری از کاربردها مقدار K در ابتدا مشخص نیست. در الگوریتم

ISODATA با در نظر گرفتن یک مقدار بزرگ برای K و با داشتن قابلیت‌های پیوند دو خوشه و

تفکیک یک خوشه، مقدار مناسب K در طی الگوریتم بدست می‌آید. مراحل الگوریتم ISODATA

به صورت زیر است [2].

مرحله 1) مقدار دهی اولیه پارامترهای مورد نیاز در الگوریتم ISODATA

(1-1) مقدار دهی اولیه یک مقدار مطلوب برای K (این مقدار به گونه‌ای مقدار دهی می‌-

شود که بدانیم تعداد خوشه‌ها از $2K$ بیشتر نیستند و همچنین از $K/2$ کمتر

نمی‌باشند)

(2-1) مقدار دهی پارامتر I که نشان‌دهنده ماکزیمم تعداد تکرار الگوریتم است.

(3-1) مقدار دهی پارامتر L که نشان‌دهنده حداکثر تعداد خوشه است که در یک تکرار

می‌تواند به هم پیوند بخورند.

(4-1) مقدار دهی پارامتر θ_N که نشان‌دهنده حداقل تعداد عضوی است که یک خوشه

می‌تواند داشته باشد (این پارامتر برای حذف کردن خوشه به کار می‌رود)

1-7 فهرست منابع

[1] P. C. Wang, M. Orr, M. Sparrow and M. Apa, "System and Method for Detecting and Deinterleaving Radar Emitters," US Patent # 7397415 B1, July 2008.

[2] P. C. Wang and C. R. Ward, "Method for Radar Pattern Recognition by Sorting Signals into Data Clusters," US Patent # 7034738 B1, April 2006.

[3] M. Ester, H. Kriegel, J. Sander and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining, KDD*, 1996.

[4] D. J. Milojevic and B. M. Popovic, "Improved Algorithm for the Deinterleaving of Radar Pulses," *IEE Proceedings*, Vol. 139, No.1, February 1992.

[5] H. K. Mardia, "New Techniques for the Deinterleaving of Repetitive Sequences," *IEE Proceedings*, Vol. 136, No.4, August 1992.

[6] Y. Kuang, Q. Shi, Q. Chen, L. Yun and K. Long, "A Simple Way to Deinterleave Repetitive Pulse Sequences," in *Proceeding of International Conference on Mathematical Methods and Computational Techniques in Electrical Engineering*, October 2005.

[7] R. J. Orsi, J. B. Moore and R. E. Mahony, "Spectrum Estimation of Interleaved Pulse Trains," *IEEE Transactions on Signal Processing*, Vol. 47, No.6, June 1999.