

شناسایی آنومالی در ترافیک دریایی بر اساس داده‌های مکانی - زمانی سیستم شناسایی خودکار (AIS)

مجتبی گودرزی، * مهدی شعبانی**

تاریخ دریافت: ۱۳۹۹/۹/۲۷ تاریخ پذیرش: ۱۳۹۹/۱۰/۱۱ نوع مقاله: پژوهشی

چکیده

نظارت بر ترافیک دریایی یکی از جنبه‌های مهم ایمنی و امنیتی به خصوص در مسیرهای دریایی شلوغ و ترافیک در نزدیکی اسکله‌هاست. در این تحقیق، مدل شناسایی آنومالی دو بخشی در ترافیک دریایی پیشنهاد شده است. در بخش اول، الگوی معمولی خط سیر و مسیر کشتی‌ها را که براساس قوانین IMO جمع‌آوری و در دسترس عموم قرار گرفته‌اند، با استفاده از الگوریتم‌های خوشه‌بندی و براساس چگالی در داده‌کاوی شناسایی شده است. مهم‌ترین حسن خوشه‌بندی بر اساس چگالی در این مرحله، شناسایی الگوی حرکتی کشتی‌ها با توجه به قوانین و شرایط موجود در منطقه دریایی است. در بخش دوم، مدل پیشنهادی با توجه به الگوی شناخته‌شده در بخش اول و با استفاده از رویکرد براساس نقطه، جابه‌جایی غیرعادی کشتی‌ها در دریا شناسایی شده است. برای خوشه‌بندی و شناسایی رفتار آنومال از طول و عرض جغرافیایی، سرعت و جهت در هر نقطه از مسیر طی‌شده به وسیله کشتی‌ها استفاده شده است. رویکرد شناسایی آنومالی براساس نقطه قابل استفاده در سیستم‌های نظارتی سیستم شناسایی خودکار (AIS) از نوع بلادرنگ نیز است. در این تحقیق، مرحله شناسایی جابه‌جایی آنومال قابلیت شخصی‌سازی از طرف تصمیم‌گیرندگان نظارتی را دارد که در نتیجه، میزان دقت شناسایی آنومالی قابل تنظیم است.

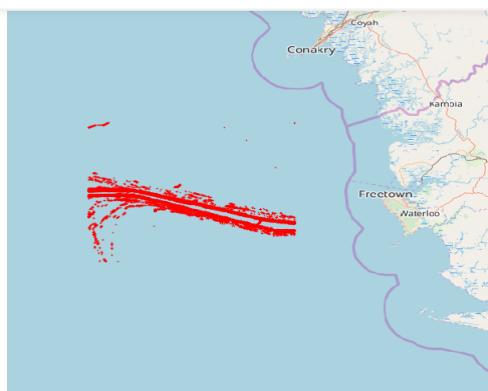
واژگان کلیدی: شناسایی آنومالی، خوشه‌بندی براساس چگالی، کاوش در داده‌های خط سیر، نظارت دریایی.

* گروه کامپیوتر، دانشکده مهندسی برق، کامپیوتر و مکانیک، دانشگاه ایوان کی، سمنان

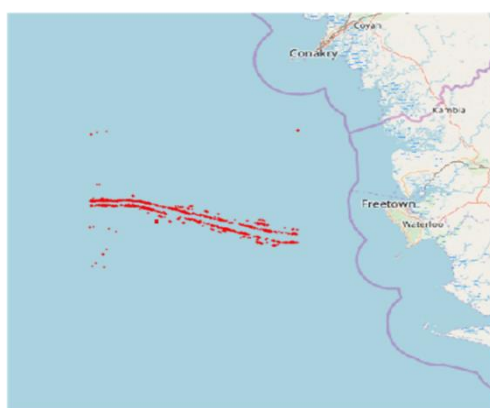
m.goodarzi1001@gmail.com

** گروه کامپیوتر، دانشکده فنی و مهندسی، دانشگاه تهران جنوب، تهران، ایران.

مجله مهندسی سیستم و بهره‌وری، سال اول، شماره ۱، اسفند ۱۳۹۹، ص ۷۳ - ۹۷



شکل ۵: مسیر رفت و آمد شناورها در پایگاه داده

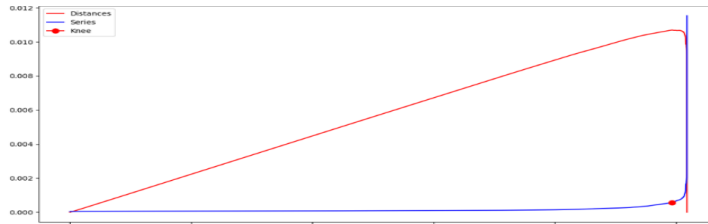


شکل ۶: مسیر رفت و آمد شناورها مربوط به قسمت
آزمون پایگاه داده یک پایگاه داده یک

۳۲۰۰۰ سطر اول از این پایگاه داده را برای آموزش و ۲۰۰۰ سطر آخر از این پایگاه داده برای آزمون استفاده شده است. نقاط موجود در پایگاه داده آزمون در هنگام آزمون برای شناسایی الگوی نرمال متحرک و ایستا استفاده نشده است. همان‌طور که در شکل ۵ مشاهده می‌شود، در این پایگاه داده شناورها در دو خط دریایی مجاور هم در حال حرکت هستند. در شکل ۶ مسیر رفت و آمد کشتی‌ها در قسمت آزمون را می‌بینید.

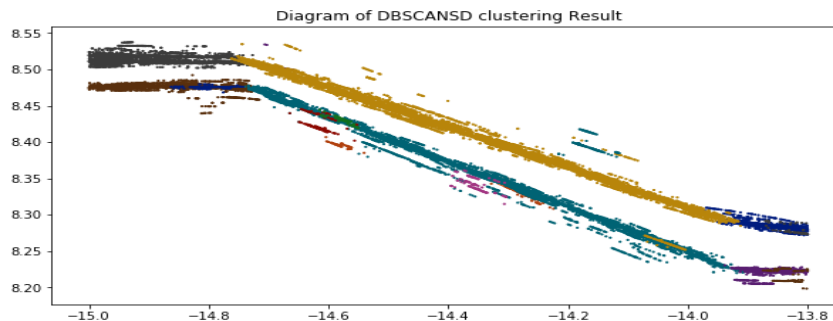
خوشه‌بندی داده‌های متحرک

تعداد حداقل نقاط همسایه^۱ برابر با ۵۰ تعیین شد. مقدار شعاع^۲ در DBSCANSD با استفاده از نمودار فاصله knn برابر با ۳۶۰۰ متر تعیین شده است. در شکل ۷ این نمودار را می‌بینید. نقطه قرمز در شکل بهترین مکان برای میزان Eps است؛ البته کمی بالاتر یا پایین‌تر از این مقدار ایرادی ندارد. از آنجا که در هنگام محاسبه فاصله‌ها از رادیان استفاده شده است، در آخر باید مقدار خروجی را در شعاع زمین ضرب کنیم تا مقدار واقعی را به متر به دست آوریم. میزان تفاوت سرعت برابر با ۱,۵ و میزان تفاوت جهت برابر با ۲,۵ تنظیم شده است.



شکل ۷: نمودار فاصله Knn برای محاسبه میزان Eps در پایگاه داده

با توجه به تنظیمات انجام‌شده، الگوریتم خوشه‌بندی DBSCANSD داده‌ها را در ۲۰ خوشه مختلف خوشه‌بندی کرده است. در شکل ۸ نتیجه این خوشه‌بندی را به صورت رنگ‌بندی شده می‌بینید. با توجه به میزان عرض GRID تعیین‌شده که برابر یک درصد تنظیم کرده‌ایم، ۲۳۴ نقطه به عنوان GV انتخاب شده‌اند. میزان عرض GRID کاملاً در تعداد این نقاط تأثیرگذار است. این مقدار به صورت تجربی به دست آمده است.

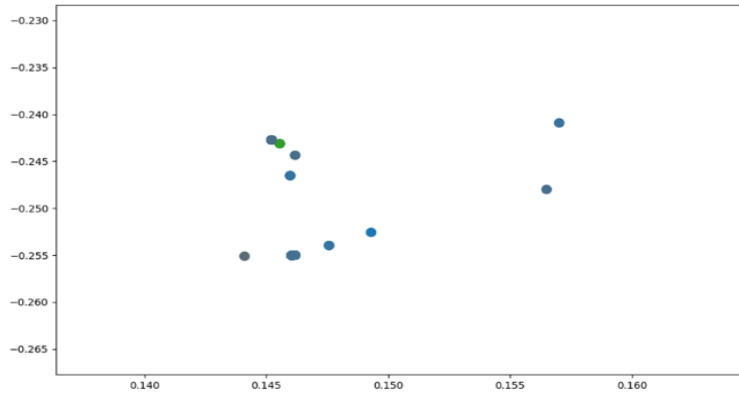


شکل ۸: تصویر رنگ‌بندی نتیجه خوشه‌بندی DBSCANSD برای نقاط متحرک قسمت آموزش

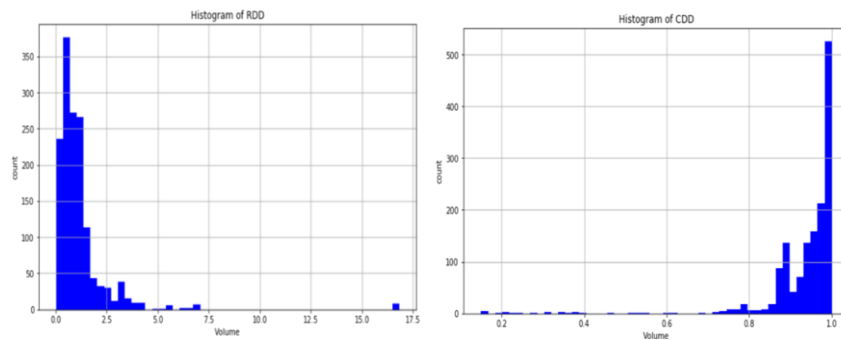
1. MinEps
2. Eps

خوشه‌بندی نقاط ایستا

برای خوشه‌بندی داده‌های ایستا با الگوریتم خوشه‌بندی HDBSCAN میزان حداقل اندازه خوشه برابر با ۵۰ تنظیم شده است. بهتر است که این مقدار برابر با حداقل نقطه همسایه^۱ در الگوریتم DBSCAN باشد. همان‌طور که در شکل ۹ می‌بینید، HDBSCAN نقاط ایستا در چهار خوشه خوشه‌بندی کرده است.



شکل ۹: نتایج خوشه‌بندی HDBSCAN برای نقاط ایستا قسمت آموزش



شکل ۱۰: نمودار هیستوگرام حداقل RDD و حداکثر CDD نقاط موجود در داده‌های آموزش

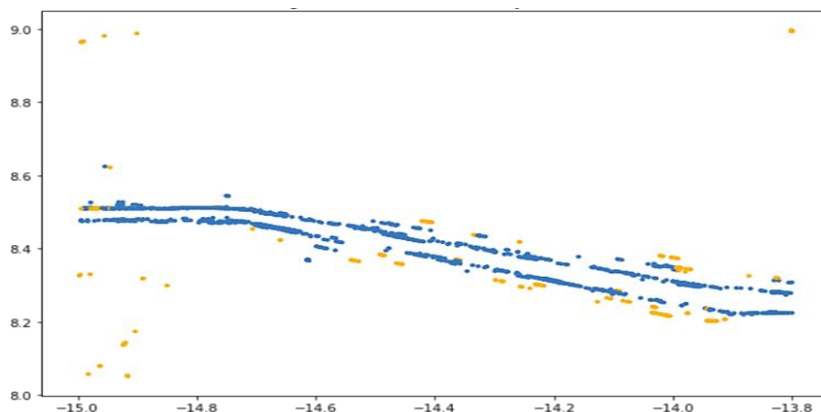
1. MinPts

نتایج شناسایی آنومالی

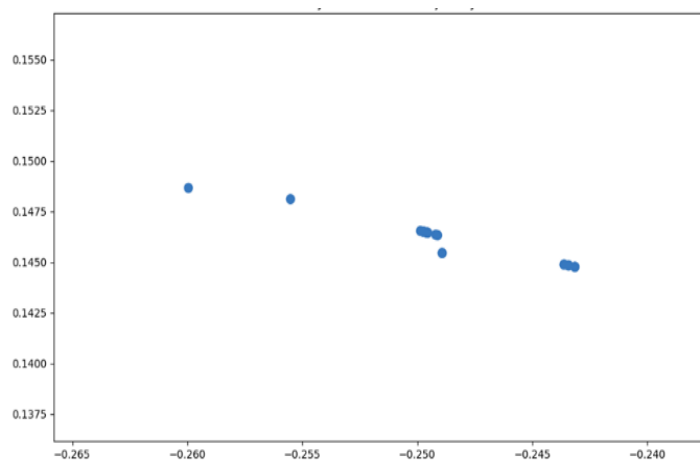
پس از بررسی نمودارهای هیستوگرام CDD و RDD، میزان آستانه CDD را برابر با ۰.۸ و میزان آستانه RDD را برابر با ۲ تنظیم شده است.

در شکل ۱۱ نتیجه خروجی میزان آنومالی در بین تمام نقاط را می‌بینید. رنگ آبی نشان نقاطی است که نرمال تشخیص داده شده‌اند.

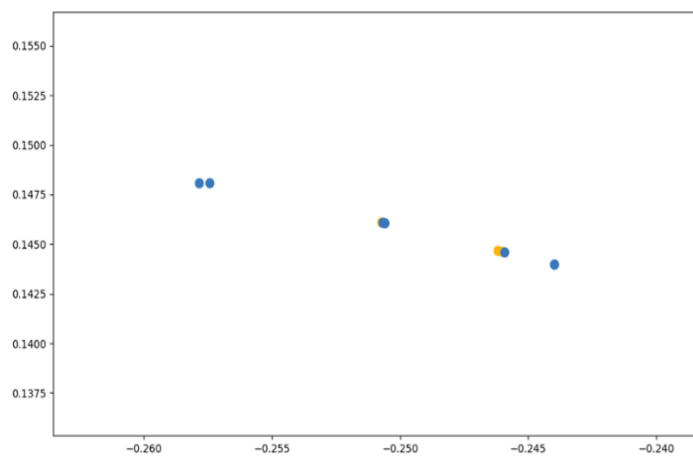
در ادامه تعدادی از خط سیرهای طی شده به وسیله کشتی‌های مختلف و میزان آنومالی تشخیص داده شده برای هر مسیر را می‌بینید. رنگ آبی نقاطی از هر خط سیر است که نرمال هستند.



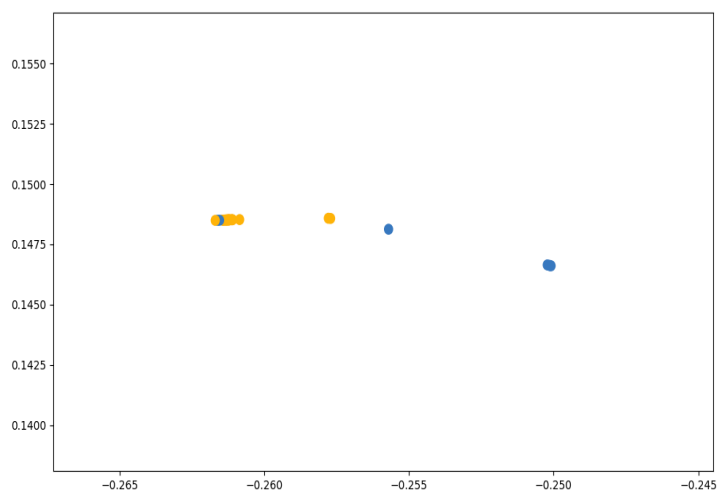
شکل ۱۱: نتایج خروجی شناسایی آنومالی در قسمت آزمون پایگاه داده



شکل ۱۲: خط سیر کشتی با شماره ۷۰۹ که کاملاً نرمال تشخیص داده شده است



شکل ۱۳: خط سیر طی شده کشتی با شماره ۵۹۴ که ۲۷ درصد آنومال تشخیص داده شده است

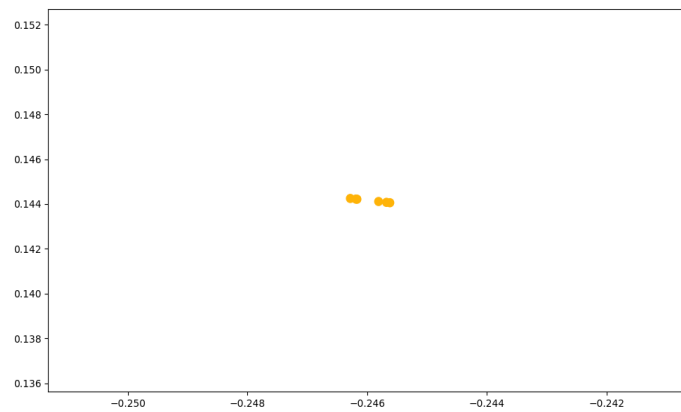


شکل ۱۴: خط سیر طی شده کشتی با شماره ۸۴۷ که دارای ۷۴ درصد آنومالی شناسایی شده است

نتیجه‌گیری

در این مقاله، ما در ابتدا با توجه به آستانه سرعت تعیین شده پایگاه داده خود را با تبعیت از مدل ایست و حرکت به دو قسمت متحرک و ایست تقسیم کرده‌ایم؛ سپس با استفاده از الگوریتم خوشه‌بندی بر اساس چگالی DBSCANSD برای نقاط متحرک و HDBSCAN برای نقاط ایستا، توانستیم الگوی نرمال حرکتی در پایگاه داده با اطلاعات مکانی-زمانی سیستم شناسایی خودکار (AIS) را شناسایی کنیم.

برای خوشه‌بندی داده‌های متحرک از طول و عرض جغرافیایی، سرعت و جهت ذخیره شده از حرکت کشتی‌ها در هر نقطه از خط سیرشان استفاده شده است. برای انتخاب پارامتر ورودی Eps در الگوریتم خوشه‌بندی DBSCANSD از نمودار فاصله Knn استفاده شده است. برای انتخاب سایر پارامترهای ورودی مورد نیاز به صورت تجربی عمل شده است. برای خوشه‌بندی و شناسایی مکان‌های ایست نرمال در محدوده دریایی پایگاه داده از طول و عرض جغرافیایی استفاده شده است.



شکل ۱۵: خط سیر طی شده کشتی با شماره ۴۸۶ که ۱۰۰ درصد انومال شناسایی شده است.

استفاده از الگوریتم خوشه‌بندی بر اساس چگالی باعث شد تا بتوانیم هر خط سیر با هر الگویی را بتوانیم خوشه‌بندی کنیم.

مدل شناسایی آنومالی در این تحقیق از نوع استاتیک و با استفاده از رویکرد براساس نقطه است. این رویکرد قابل استفاده در سیستم‌های شناسایی آنومالی به صورت بلادرنگ است. از دو فاصله RDD و CDD در نقاط متحرک به عنوان معیار طبیعی (نرمال) یا خارج از محدوده بودن هر نقطه از نقاط دارای سرعت بالاتر از آستانه تعیین شده استفاده شده است. برای تنظیم مقدار این دو پارامتر

از نمودار هیستوگرام حداقل RDD و حداکثر CDD نقاط موجود در داده‌های آزمون استفاده شده است. برای شناسایی نقاط ایست خارج از محدوده از الگوریتم شناسایی خارج از محدوده GLOSH استفاده شده است. تنظیم‌پذیر بودن RDD و CDD برای داده‌های متحرک سبب شده است که میزان دقت مدل پیشنهادی قابل تنظیم باشد که این امر باعث کاربردی بودن این تحقیق در دنیای واقعی می‌شود. استفاده از محاسبات توزیع‌شده یکی از موضوعاتی است که می‌تواند باعث افزایش سرعت الگوریتم‌های خوشه‌بندی در هنگام به دست آوردن الگوی حرکتی نرمال شود.

منابع

- Abbasifard, M. R., & naderi, hassan. (2017). Indexing the past and current position of moving objects in large-scale dataset. *Journal of Geospatial Information Technology*, 5(2), 141–162. <https://doi.org/10.29252/jgit.5.2.141>
- Ball, H., & Satellite, A. I. S. (2013). *Satellite AIS for Dummies*. Mississauga, ON: Wiley.
- Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 160–172). Springer.
- Campello, R. J. G. B., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 10(1), 5.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 15.
- Daszykowski, M., Walczak, B., Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, pp. 226–231). <https://doi.org/10.1016/B978-044452701-1.00067-3>
- De Vries, G. K. D., & Van Someren, M. (2012). Machine learning for vessel trajectories using compression, alignments and domain knowledge. *Expert Systems with Applications*, 39(18), 13426–13439.
- Elbatta, M. T. H., & Ashour, W. M. (2013). A dynamic method for discovering density varied clusters. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(1), 123–134.
- IMO. (2018). About IMO. <https://doi.org/10.1007/bf02532644>
- Kassambara, A. (2018). DBSCAN: Density-Based Clustering Essentials - Datanovia. Retrieved February 15, 2019, from <https://www.datanovia.com/en/lessons/dbscan-density-based-clustering-essentials/>
- Kazemi, S., Abghari, S., Lavesson, N., Johnson, H., & Ryman, P. (2013). Open data for anomaly detection in maritime surveillance. *Expert Systems with Applications*, 40(14), 5719–5729.
- Lane, R. O., Nevell, D. A., Hayward, S. D., & Beaney, T. W. (2010). Maritime anomaly detection and threat assessment. In *Information Fusion (FUSION), 2010 13th Conference on* (pp. 1–8). IEEE.
- Laxhammar, R. (2008). Anomaly detection for sea surveillance. In *Information Fusion, 2008 11th International Conference on* (pp. 1–8). IEEE.
- Liu, B., de Souza, E. N., Matwin, S., & Sydow, M. (2014). Knowledge-based clustering of ship trajectories using density-based approach. In *2014 IEEE International Conference on Big Data (Big Data)* (pp. 603–608). IEEE. <https://doi.org/10.1109/BigData.2014.7004281>
- Liu, B., Souza, E. N. De, Hilliard, C., & Matwin, S. (2015). Ship Movement

- Anomaly Detection Using Specialized Distance Measures. *2015 18th International Conference on Information Fusion, Fusion 2015*, 1113–1120. Retrieved from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84960498152&partnerID=40&md5=5ec29d27f2707d1e546d7d47d68b008a>
- Mazimpaka, J. D., & Timpf, S. (2016). Trajectory data mining: A review of methods and applications. *Journal of Spatial Information Science*, *2016*(13), 61–99.
- McInnes, L., & Healy, J. (2017). Accelerated Hierarchical Density Based Clustering. In *Data Mining Workshops (ICDMW), 2017 IEEE International Conference on* (pp. 33–42). IEEE.
- Mehrotra, K. G., Mohan, C. K., & Huang, H. (2017). *Anomaly Detection Principles and Algorithms*. Springer.
- Morris, B. T., & Trivedi, M. M. (2008). A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, *18*(8), 1114.
- Pallotta, G., Vespe, M., & Bryan, K. (2013). Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. *Entropy*, *15*(6), 2218–2245.
- Palma, A. T., Bogorny, V., Kuijpers, B., & Alvares, L. O. (2008). A clustering-based approach for discovering interesting places in trajectories. In *Proceedings of the 2008 ACM symposium on Applied computing* (pp. 863–868). ACM.
- Rahmah, N., & Sitanggang, I. S. (2016). Determination of Optimal Epsilon (Eps) Value on DBSCAN Algorithm to Clustering Data on Peatland Hotspots in Sumatra. *IOP Conference Series: Earth and Environmental Science*, *31*(1). <https://doi.org/10.1088/1755-1315/31/1/012012>
- Rocha, J. A. M. R., Times, V. C., Oliveira, G., Alvares, L. O., & Bogorny, V. (2010). DB-SMoT: A direction-based spatio-temporal clustering method. In *Intelligent systems (IS), 2010 5th IEEE international conference* (pp. 114–119). IEEE.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.
- Satopaa, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops* (pp. 166–171). IEEE. <https://doi.org/10.1109/ICDCSW.2011.20>
- Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., & Vangenot, C. (2008). A conceptual view on trajectories. *Data & Knowledge Engineering*, *65*(1), 126–146.
- Van Brummelen, G. (2013). *Heavenly mathematics: the forgotten art of spherical trigonometry*. Princeton University Press. Retrieved from

<https://books.google.com/books?id=0BCCz8Sx5wkC&pg=PR7>

Veness, C. (2002). Calculate Distance and Bearing between Two Latitude/Longitude Points Using Haversine Formula in JavaScript. *MIT Open Source*, 1–39. Retrieved from <https://www.movable-type.co.uk/scripts/latlong.html>

Vespe, M., Visentini, I., Bryan, K., & Braca, P. (2012). Unsupervised learning of maritime traffic patterns for anomaly detection.