

국내 연안 해역 선박 항적 군집화를 위한 항적 간 거리 척도 개발 연구

이승주* · 이원희** · 민지홍*** · 조득재**** · † 박현우

*,***서울대학교 석사과정생, **선박해양플랜트연구소 선임연구원,
****선박해양플랜트연구소 책임연구원, † 서울대학교 데이터사이언스대학원 교수

Research on the Development of Distance Metrics for the Clustering of Vessel Trajectories in Korean Coastal Waters

Seungju Lee* · Wonhee Lee** · Ji Hong Min*** · Deuk Jae Cho**** · † Hyunwoo Park

*,***Student, Graduate School of Data Science, Seoul National University. Gwanak-gu, Seoul, Korea

**Senior Researcher, KRISO, 32 Yuseong-daero, Yuseong-gu, Seoul, Korea

****Principal Researcher, KRISO, 32 Yuseong-daero, Yuseong-gu, Seoul, Korea

† Professor, Graduate School of Data Science, Seoul National University. Gwanak-gu, Seoul, Korea

요 약 : 본 연구에서는 국내 연안 해역 환경에서의 해상교통관제 서비스에 기여할 수 있는 항적 간 거리 척도를 개발하였다. 새로운 항적 간 거리 척도는 전통적으로 위치 시계열 간의 유사도를 측정하는 데 활용되는 하우스도르프 거리(hausdorff distance)와 두 항적 간의 대지속력(Speed Over Ground, SOG)의 평균 간의 차이, 그리고 대지침로(Course Over Ground)의 분산 간의 차이를 가중합하여 설계되었다. 새로운 척도의 유효성을 검증하기 위하여 실제 AIS 항적 데이터와 병합 군집화 알고리즘을 활용한 기존 항적 간 거리 척도와의 비교 분석이 수행되었으며, 새로운 거리 척도를 활용한 항적 군집화 결과가 하우스도르프 거리(hausdorff distance), 그리고 다이내믹 타임 워핑 거리(Dynamic Time Warping distance) 등 기존 척도에 비해 항적 간 지리적 거리나 대지속도 및 대지침로 등 선박 거동 특성의 분포를 비슷하거나 그 이상의 수준으로 정교하게 반영하고 있음을 데이터 시각화로서 확인하였다. 정량적으로는 Davies-Bouldin 지표를 기준으로, 군집화 결과가 더욱 우수하거나 약간 낮은 수준을 기록한 한편, 거리 계산 효율성에서는 특히 우수함을 실증하였다.

핵심용어 : 다이내믹 타임 워핑 거리, 병합 군집화 알고리즘, 선박자동식별시스템(AIS), 하우스도르프 거리, 해양안전

Abstract: This study developed a new distance metric for vessel trajectories, applicable to marine traffic control services in the Korean coastal waters. The proposed metric is designed through the weighted summation of the traditional Hausdorff distance, which measures the similarity between spatiotemporal data and incorporates the differences in the average Speed Over Ground (SOG) and the variance in Course Over Ground (COG) between two trajectories. To validate the effectiveness of this new metric, a comparative analysis was conducted using the actual Automatic Identification System (AIS) trajectory data, in conjunction with an agglomerative clustering algorithm. Data visualizations were used to confirm that the results of trajectory clustering, with the new metric, reflect geographical distances and the distribution of vessel behavioral characteristics more accurately, than conventional metrics such as the Hausdorff distance and Dynamic Time Warping distance. Quantitatively, based on the Davies-Bouldin index, the clustering results were found to be superior or comparable and demonstrated exceptional efficiency in computational distance calculation.

Keywords: Dynamic Time Warping distance, agglomerative clustering, AIS, Hausdorff distance, marine safety

1. 서 론

전세계 물동량에서 해상 운송이 담당하는 비율이 90%에 육박함에 따라 해역 혼잡에 따른 해양 안전 위협을 경감하기 위하여 해운 분야에서는 선박자동식별시스템 (AIS, Automatic Identification System)으로 송신 및 수신되는 선박 항적 정보를 활용한 방법론 개발에 적극적으로 대응하고

있다(Zhao and Shi, 2019; Zhen et al., 2017). 특히 대표적인 머신러닝 기법인 항적 군집화 방법론을 기반으로 유사한 선박들의 움직임, 혹은 다수 해상 교통에서의 유의미한 패턴을 포착하여 선박 운항 감시 및 안전 관제 서비스 등에 활용하기 위한 연구가 활발하다(Zhao and Shi, 2019).

특히 항구 인근의 해역은 항만 및 물류 시설에 대한 수요로 인해 일반적으로 혼잡도가 높으며(Oh et al., 2018),

† Corresponding author : 정회원, hyunwoopark@snu.ac.kr 02)880-1139

* 정회원, lotusist10@snu.ac.kr

** 정회원, weelon@kriso.re.kr 042)866-3718

*** 정회원, jhmin@snu.ac.kr

**** 정회원, djcho@kriso.re.kr 042)866-3683

항구의 출입 통로가 되는 국내 연안, 특히 해안선이 복잡하여 해역이 좁은 서남해안에서는 혼잡도 증가에 따른 충돌, 좌초 등의 해사 안전 리스크에 더욱 취약할 수밖에 없다(Lee et al., 2010). 특정 해역에서의 높은 교통 밀집도는 선박 충돌 등 직접적인 안전 위협을 야기할 뿐 아니라, 해상교통관제 업무 부하를 높여 원활한 관제 서비스를 저해할 수 있다(Oh et al., 2018; Park and Park, 2022).

선박 항적 데이터는 무선통신 시스템을 통해 실시간으로 수집되는 대용량 위치 시계열 데이터이다. 그리고 선박 항적 군집화는 유사한 항적들을 하나의 군집으로, 이질적인 항적들을 다른 군집으로 판별하는 과정이다. 다수의 선박 항적들을 소수의 군집으로 요약하여 항적 데이터의 양과 복잡성을 줄인다면, 관제 서비스 업무 부하를 크게 경감시키고, 선박 안전 운항 리스크 또한 줄여 국내 연안 선박의 항행 안전 제고에 기여할 수 있을 것이다. 특히 항만 인근을 비롯한 국내의 좁은 연안 해역 환경에서의 해운 교통량이 증가하는 추세에서, 관제 업무 부하의 경감 효과는 더욱 클 것이다.

다만 실제 해상 안전에 기여할 수 있는 유효한 항적 군집화를 위해서는 항적 및 선박의 특성을 실제적으로 반영한 유사도 측정 척도가 선제되어야 한다. 이때 단순히 지리적 위치의 가까움만으로 항적 간 유사도가 측정되기보다는, 비슷한 해역이어도 출입항 목적, 선종, 해안선 또는 섬과의 거리 등에 따른 다양한 동적 특성의 유사성도 반영되어야 할 것이다. 항적 정보의 높은 정보량과 복잡성을 몇 개의 항적 군집, 또는 군집 내 대표 항적으로써 줄이되, 항적 데이터의 지리적 및 동적 특성을 포괄하여 유효하게 표현하는 항적 간 유사도 척도가 군집화 과정의 기준이 된다면, 관제 서비스의 부담을 줄이면서도 선박 특성에 맞는 안전한 항로를 제공하는 데 기여할 수 있을 것이다.

본 논문에서는 항적 데이터와 관련된 거리 측정 방법론과 군집화 알고리즘 등에 관련된 기존 연구들을 우선 살피고, 그 성과들을 참고하여 두 항적 간의 지리적 거리뿐 아니라 대지속력(SOG, Speed Over Ground)의 평균 간의 차이와 대지침로(COG, Course Over Ground)의 분산 간의 차이를 활용하여 선박 항적의 동적 특성들을 포괄적으로 반영하는 새로운 항적 간 거리 척도를 제안하였다. 이어서 개발한 신규 거리 척도가 국내 연안 해역에서 유효하게 활용될 수 있는지를 실증하기 위하여 국내 주요 항만을 출입한 실제 선박 항적 데이터를 기반으로 항적 간 거리 계산 및 항적 군집화 과정을 수행하였다. 그리고 기존 항적 간 거리 계산 척도와 비교하여 개발된 척도의 가치를 평가하기 위하여 시각적 및 정량적 방법론으로 다양하게 분석하였다.

2. 관련 연구

2.1 항적 간 거리 척도

군집화 대상인 선박의 항적은 시간의 흐름에 따라 상태가 변하는 동적인 객체의 정보를 담고 있다는 특성이 있어 단일한 척도로써 그 유사도를 측정하기가 어렵다. 이에 한 쌍의 항적 간의 유사도, 혹은 그 반대 척도인 거리를 하나의 실수로 표현하기 위한 다양한 측정 방법론이 개발되어 왔다(Buchin et al., 2012; Magdy et al., 2015).

선박 항적 W 는 특정 시점의 선박의 위도 Lat 및 경도 Lon 좌표로 표현되는 위치 정보, 그리고 대지침로 COG 및 대지속력 SOG 등 거동 정보로 표현되는 웨이포인트 waypoints w 의 집합이다. 이는 식 (1)과 같이 정의할 수 있다.

$$W = \{w_0, w_1, \dots, w_n\} \quad (1)$$

where $w_t = [Lat_t, Lon_t, SOG_t, COG_t]$

즉, 선박의 항적 데이터는 기록된 시점에 따라 변화한 선박의 위치 및 동적 정보의 추이를 저장하고 있는 위치 시계열 데이터이다. 일련의 웨이포인트로 구성된 항적의 기하학적 특성을 반영하거나, 혹은 시계열적 특성을 반영함으로써 항적 간 거리는 다음과 같이 여러 방식으로 정의될 수 있다.

1) 하우스도르프 거리(Hausdorff distance)

하우스도르프 거리는 여러 꼭지점으로 이루어진 서로 다른 두 집합의 거리를 측정하기 위하여 활용되는 방법론으로서, 두 항적의 기하학적 유사도를 측정한다(Alt, 2009). 하우스도르프 거리를 구하기 위해서는 우선 서로 다른 항적을 구성하는 두 개의 웨이포인트들을 끝점으로 하는 최소 거리들을 모두 계산한다. 하우스도르프 거리는 그 최소 거리들 중의 최댓값이다. 각각 웨이포인트 a, b 의 집합으로 이루어진 항적 A, B 가 있을 때 두 항적의 하우스도르프 거리 $hausdorff(A, B)$ 를 수식으로 정의한 것이 식 (2)이다.

$$hausdorff(A, B) = \max(h(A, B), h(B, A)) \quad (2)$$

where $h(A, B) = \max_{a \in A} (\min_{b \in B} | | a - b | |)$

2) 다이내믹 타임 워핑 거리(Dynamic Time Warping distance, DTW)

다이내믹 타임 워핑 거리도 서로 다른 두 개의 시계열 간의 거리를 측정하는 데 널리 사용되는 방식으로, 하우스도르프 거리 척도와 유사하게 서로 다른 항적 상에

존재하는 두 개의 웨이포인트로 가능한 모든 거리 조합을 재귀적으로 계산한다(Keogh and Ratanamahatana, 2005). 이때 두 웨이포인트의 시점상 순서가 일치하지 않더라도 웨이포인트 간의 최적의 거리를 탐색하는 연산이 수행되므로, 서로 다른 길이의 시계열 데이터 간의 유사도를 측정할 수 있다는 장점이 있으나, 연산 상의 부담이 비교적 크다는 문제가 있다(Magdy et al., 2015).

각각 웨이포인트 a, b 로 이루어진 항적 A, B 의 다이내믹 타임 워핑 거리 $DTW(A, B)$ 를 구하는 식은 식 (3)과 같다. 이때 $Rest(A)$ 는 첫 번째 웨이포인트를 제외한 항적 A 의 항적 데이터를 의미한다(Magdy et al., 2015).

$$DTW(A, B) = \begin{cases} 0 & \text{if } m = n = 0 \\ \infty & \text{if } m = 0 \text{ or } n = 0 \\ dist(a_1, b_1) + \min \begin{cases} DTW(Rest(A), Rest(B)) \\ DTW(Rest(A), B) \\ DTW(A, Rest(B)) \end{cases} & \text{otherwise} \end{cases} \quad (3)$$

3) 개선된 하우스도르프 거리

하우스도르프 거리는 단순히 웨이포인트의 위도 및 경도 정보만을 활용하여 계산된 지리적 거리 척도이므로, 선박의 대지속력 및 대지침로 정보는 반영되지 않는다. 이에 Zhen et al.(2017)은 두 항적 간의 하우스도르프 거리, 그리고 대지침로값의 차이를 가중합으로 결합하여 새로운 거리 척도를 고안하였다.

본 연구에서는 다이내믹 타임 워핑 거리에 비해 하우스도르프 거리가 연산 측면에서 부하가 적다는 점, 그리고 항적 간 지리적 거리뿐 아니라 선박의 거동 정보를 활용하여 항적의 유사도를 평가할 수 있다는 점에 착안하여, Zhen et al.(2017)의 척도를 발전시킨 형태의 항적 간 거리 비교 방법론을 개발하였다. 본 연구에서 제안하는 항적 간 거리 측정법은 항적 간의 대지침로의 평균 간 차이가 아닌 분산의 차이를 반영할 뿐만 아니라, 대지속도의 평균 간 차이까지 반영하여 선박의 거동을 종합적으로 반영하는 방법론이다.

2.2 군집화 알고리즘

앞서 언급하였듯이 항적 간 거리 또는 유사도는 활용하는 척도에 근거하여 다양하게 정의될 수 있으며, 군집화(Clustering) 모델은 바로 이 거리 척도에 근거하여 유사한 개체들을 동일 군집으로 평가한다. 군집화 알고리즘은 공통적으로 해사 안전과 관련된 선행 연구들에서는 다수의 항적을 군집으로 만들어 그 정보를 요약하기 위한 목적으로 사용되었으며, 활용된 알고리즘은 목표 군집 개수인 K 개의 군집 중심점을 갱신해가는 형식의 K -means 알고리즘을 활용하거나(Kim et al.,2014; Oh et al.,2018) 밀도 기반의 DBSCAN, HDBSCAN 등이 주로

활용되었다(Rong et al.,2020; Dobrkovic et al.,2018; Pallotta et al.,2013).

3. 방법론

3.1 새로운 항적 간 거리 척도

본 연구에서는 지리적으로 복잡한 국내 연안에서 유효하게 활용될 수 있는 항적 간 거리 측정 방법론을 개발하고자 한다. 이에 항적 간 지리적 거리뿐 아니라 항적의 형태, 선박의 동적 특성 등을 포괄적으로 반영하고자 식 (4)와 같이 두 항적 간의 대지속력의 평균 간의 차이, 그리고 대지침로의 분산 간의 차이를 하우스도르프 거리와 가중합한 산식을 설계하였다.

$$D_{new} = k_1 * hausdorff(A, B) + k_2 * (| SOG_{mean}^A - SOG_{mean}^B |) + k_3 * (| COG_{var}^A - COG_{var}^B |)$$

$$where SOG_{mean}^A = \frac{1}{N_A} \sum_{i \in A} SOG_i^A$$

$$where COG_{var}^A = \frac{1}{N_A} \sum_{i \in A} (COG_i^A - COG_{mean}^A)^2$$

$$where COG_{mean}^A = \frac{1}{N_A} \sum_{i \in A} COG_i^A$$

$$subject to k_1 + k_2 + k_3 = 1 \quad (4)$$

두 항적의 대지속력의 평균 간의 차이는 각 항적의 선박 간의 전반적인 속도 차이를 의미하며, 이는 선박 대표적인 동적 특성에 해당한다. 한편 두 항적의 대지침로의 분산 간의 차이는 두 항적의 직진성의 정도, 즉 항적의 형태적 차이를 반영하기 위한 항이다. 가령 방향의 변화가 적거나 직진하는 항적은 대지침로의 분산이 작을 것이며, 변화가 큰 항적은 분산이 크다는 것을 감안한 것이다.

새롭게 정의된 항적 간 거리 측정 척도는 항적 간 지리적 거리와 선박의 동적 특성, 그리고 항적 형태를 종합적으로 반영하기 위하여 설계된 것이다. 이 거리 척도를 적용하여 군집화 대상인 항적들 전체에 대하여 서로 간의 거리값들을 계산하면, 이 모든 거리 정보를 거리 매트릭스(distance matrix)에 저장할 수 있다. 거리 매트릭스의 각 행 또는 열은 각 항적에 대응되므로 이는 모든 대각 원소의 값이 0인 대칭 행렬(symmetric matrix)이다. 거리 매트릭스에는 모든 항적 간 거리 정보가 저장되어 있으므로 이를 서로 거리가 적음, 즉 유사한 항적끼리 군집화하는 알고리즘에 활용된다.

3.2 병합 군집 알고리즘(Agglomerative Clustering)

병합 군집 알고리즘은 미리 목표한 군집 개수가 달성될 때까지 유사한 군집들이 점진적으로 더 큰 군집을 형성하는 알고리즘이다(Fisher, 1996).

알고리즘의 초기 설정에서는 군집화 대상, 즉 전체 선박 항적들이 모두 서로 다른 군집으로 간주된다. 더 큰 군집을 형성하기 군집 간 유사도는, 앞서 3.1.에서 언급된 항적 전체에 대한 거리 매트릭스를 활용하여 계산된다. 군집화 진행 중에는 군집 간의 거리, 또는 유사도를 군집 내 항적들의 거리에 대한 평균값으로 평가하도록 설정되며, 군집의 개수가 목표 군집 개수보다 많을 경우, 가장 유사한 군집들을 합쳐가며 알고리즘의 최종 단계에서는 목표한 군집 개수에 따라 각 항적들이 각각의 군집에 속하게 된다.

본 연구에서는 새로운 항적 간 거리 척도의 유효성을 분석하기 위하여 병합 군집 알고리즘을 분석 대상 항적들에 적용한 결과를 활용하였다. 새로운 척도를 기존 척도와 비교하여야 하므로 동일한 목표 군집 개수를 설정하였으며, 군집 내 또는 군집 간 특성을 확인하여 균질한 특성의 항적들이 동일 군집으로 포착되었는지, 혹은 이질적인 항적들이 각각 다른 군집으로서 평가되었는지를 다양한 방식으로 분석하였다.

4. 분석 대상 데이터

본 연구에서는 AIS로 수집된 선박 항적 데이터를 활용하여 국내 해역에서의 선박 항적 군집화에 적합한 거리 측정 방법론을 개발하고 검증하였다. 식 (1)에서 정의하였듯이 항적 데이터는 AIS 통신 시스템으로 실시간으로 수집되는 위도 및 경도, 대지침로와 대지속력 등의 정보를 담고 있으며, 각 선박에 대한 식별 정보는 암호화되어 있다. 다만 동일한 선박 식별자라면 암호화된 식별 정보도 동일한 값을 가지기 때문에 각 웨이포인트들이 동일 선박의 항적인지를 판단하는 것에는 문제가 없었다. 그 밖의 구체적인 실험 환경은 이후 자세한 실험 및 고찰 부분에서 다룬다.

항행 중인 선박의 항적들을 대상으로 항적 간 거리를 측정하고 군집화를 진행하기 위하여, 위도 및 경도 값에서 이상치가 있는 항적은 제외하였다. 아울러 최대 속력이 2knots(3.7km/h) 이하인 항적은 정지한 선박의 항적 데이터로 간주하여 마찬가지로 모두 분석 대상 데이터에서 제외하였다.

식 (4)와 같이 위치 정보와 선박 동적 정보를 결합하여 정의된 새로운 거리 척도를 실제 항적 데이터에 적용하려면 각 특성의 범위 및 단위 차이를 고려하여 정규화하여야 한다. 대상 항적 데이터의 위도 및 경도, 대지속력에 대해서는 최솟값을 0, 최댓값을 1로 매핑하여 정규화하였으며, 각도인 대지침로는 사인(sine) 함수를 활용하여 마찬가지로 0 과 1 사이의 값으로 정규화하였다.

5. 실험 및 고찰

본 연구에서는 개발한 새로운 거리 척도의 가치를 검증하기 위하여, 신규 거리 척도 및 기존의 거리 척도들 각각의 방법으로 계산된 거리 매트릭스에 동일한 병합 군집화 알고리즘을 적용하여 그 결과를 비교하는 실험을 수행하였다.

앞서 3.1에서 서술하였듯이, 본 연구에서는 항적의 형태와 선박의 동적 특성의 차이, 그리고 항적 간 지리적 거리를 반영한 거리 척도를 고안하였다. 이 거리는 식 (4)의 각 항에 부여되는 가중치 조합에 의해 다양하게 정의될 수 있다. 항적 군집화 과정은 분석 대상 전체 항적들 간에 가능한 모든 거리에 대한 값을 저장한 거리 매트릭스에 기반하여 이루어진다. 즉, 항적의 개수가 많아질수록 거리 매트릭스를 작성하는 데 드는 연산량은 제곱으로 증가하며, 대상 항적 개수가 동일하더라도 정의되는 항적 간 거리 척도가 늘어나면 계산하여야 하는 거리 매트릭스의 수도 그에 비례하여 늘어날 것이다. 그러므로 기존 항적 간 거리 척도와 새로운 거리 척도를 비교하는 실험 이전에, 새로운 거리 척도를 하나로 정의할 수 있는 최적의 가중치 조합을 확인하기 위한 실험이 선행될 것이다.

기존 항적 간 거리 척도와 비교하여 새로운 거리 척도의 유효성을 검증하는 실험에서는 실제 항적 데이터에 대해 하우스도르프 거리와 다이내믹 타임 워핑 거리, 그리고 선행된 실험에서 발견된 최적의 가중치 조합으로 정의된 새로운 거리 등 총 세 가지 거리 척도에 기반한 군집화 결과를 다양한 측면에서 비교하였다. 우선 각 척도에 기반한 거리 매트릭스가 얼마나 신속하게 계산되는지를 비교함으로써 연산 측면에서의 효율성을 비교하였다. 이어서 세 가지 거리 매트릭스에 동일 군집화 알고리즘을 적용한 결과를 정량적 척도, 그리고 군집별 항적 데이터의 특성별 분포를 확인할 수 있는 시각화 방법론을 활용하여 비교하였다.

항적 병합 군집화 알고리즘의 조건은 모든 실험에서 동일하게 적용되었다. 최종 군집 개수가 5개가 될 때까지 유사한 군집을 병합하도록 설정되었으며, 군집간 유사도는 서로 다른 군집에 속하는 두 항적 간 거리값들의 평균값으로 계산되었다.

군집화 결과에 대한 정량적 평가는 군집 내 변동 및 군집 간 유사도를 직관적으로 측정하는 Davies-Bouldin 지표(DB)를 활용하며, 이는 총 N 개의 군집에 대해 군집 i 와 군집 j 에 대한 중심점(centroids) c 사이의 유클리드 거리(Euclidean distance)로 정의된 군집 간 거리 M 을 분모로 하고, 군집 내 중심점 대비 항적 웨이포인트 X 들의 변동성으로 정의된 S 를 분자로 하는 지표 R 로써 식 (5)와 같이 정의된다(Davies and Bouldin, 1979). 즉, 군집 내 변동성은

작고, 군집 간 거리는 멀도록 잘 군집화되어 있으면 Davies-Bouldin 지표의 값은 작아진다.

$$DB = \frac{1}{N} \sum_{i=1}^N R_i$$

where $R_i \equiv \max_{i \neq j} R_{ij}$

where $R_{ij} = \frac{S_i + S_j}{M_{ij}}$

where $S_i = \frac{1}{N_i} \sum_{j=1}^{N_i} ||X_j - c_i||$

where $M_{ij} = ||c_i - c_j|| = \sqrt{\sum_{k=1}^N ||a_{ki} - a_{kj}||}$

섬이 많고 해안선이 복잡한 국내 연안 환경에서의 유효성을 실증하기 위하여, 2022년 11월 간 서남해안의 주요 항구에 입항하는 비어선 선박들의 실제 항적들이 실험에서 활용되었다.

5.1 새로운 항적 간 거리 척도의 최적 가중치 조합 탐색

새로운 항적 간 거리를 정의할 최적의 가중치 조합을 탐색하기 위하여, 2022년 11월 간 포항항에서 출항하여 광양항에 입항한 비어선 선박들의 37개의 항적들을 대상으로, 세 가지의 가중치 조합으로 계산한 항적 간 거리 매트릭스를 기반으로 군집화를 수행하였다. 가중치 조합 1은 하우스도르프 거리값이 있는 항에 가장 큰 가중치인 0.5를 두고, 그 외의 항에서는 0.25로 균등하게 가중치를 부여한 조합이다. 비슷한 방식으로 조합 2는 선박 간 평균 대지속력의 차이, 그리고 조합 3은 대지침로의 분산에 가장 큰 가중치를 부여하였다.

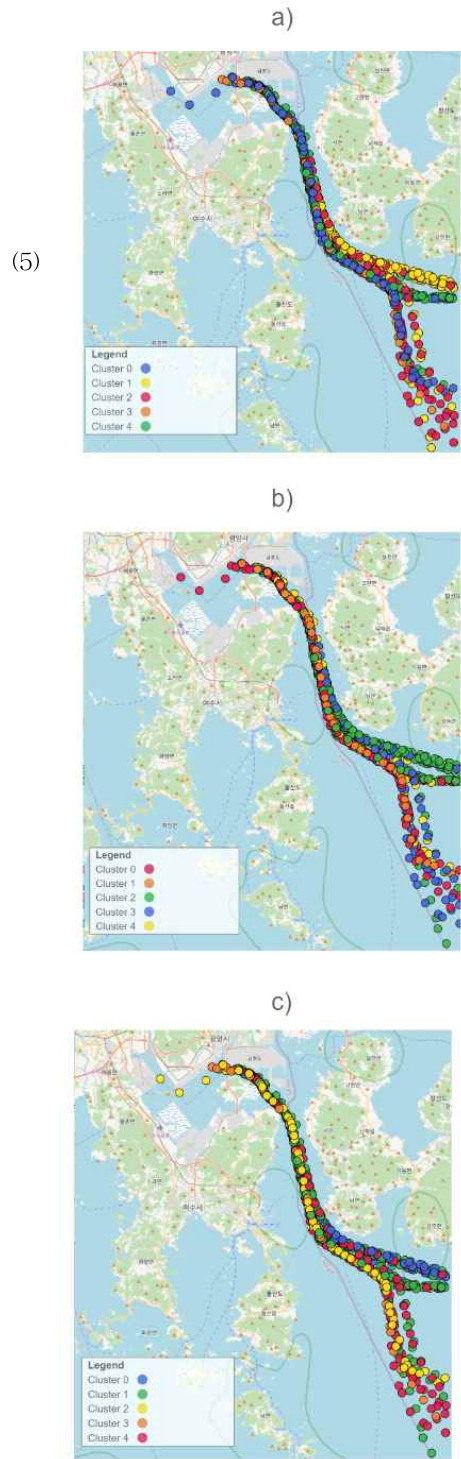


Fig. 1 Clustering results for different weight combinations in defining a new distance between trajectories.
 a) Weight Combination 1
 b) Weight Combination 2
 c) Weight Combination 3

Fig. 1의 a), b), 그리고 c)는 각각 가중치 조합 1, 조합 2, 그리고 조합 3을 적용하여 항적들을 다섯 개의 군집으로

만들어 같은 군집은 동일한 색으로 항적의 웨이포인트들을 지도상에 표시한 결과이다. 지리적 정보와 항적의 형태를 기준으로 광양항으로 접근하는 항적들은 섬과 비교적 가까운 항적, 그리고 먼 바다로부터 접근하는 항적 크게 두 가지로 분기되며, 세 가지 조합의 군집화 결과가 전반적으로 그러한 항적의 형태 차이를 반영하고 있음을 확인할 수 있다.

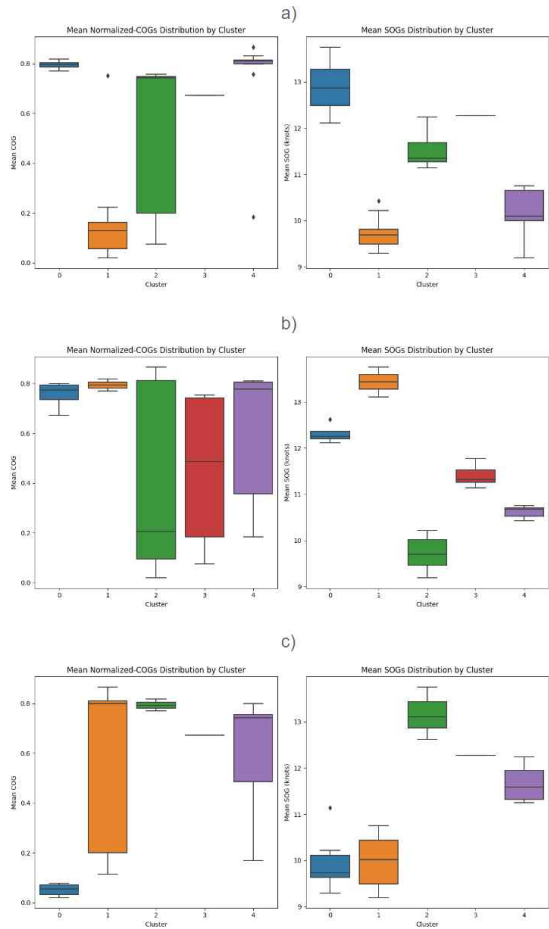


Fig. 2 Distributions of mean normalized COG values and mean SOG values in clustering results for different weight combinations in defining a new distance between trajectories. a) Weight Combination 1 b) Weight Combination 2 c) Weight Combination 3

항적의 형태나 위치 정보 외에도 선박의 거동 관련 정보가 군집화에 반영되는지 분석하기 위하여 각 군집에 속하는 항적들의 평균 대지침로값과 평균 대지속력의 분포를 확인하였다. Fig. 2의 a), b), 그리고 c)는 각각 가중치 조합 1, 조합 2, 그리고 조합 3으로 항적 간 거리를 정의하고, 각 군집 내 항적들 각각의 정규화된 대지침로값의 평균값들, 그리고 대지속력의 평균값들 분포를 박스 플롯(box plot)으로 시각화한 것이다. 각각의 시각화 자료에서, 박스의 높이가 짧을수록 해당 군집에서는 평균으로 대표되는 선박 거동의 경향성이 비교적 균질한 항적들이 묶였다고 해석할

수 있다. 또, 각 군집의 박스의 높이가 각각 다른 범위에 걸쳐 있다면, 해당 군집화 결과는 이질적인 특성을 지닌 항적들을 다른 군집으로서 잘 분류하였다고 해석할 수 있다.

Table 1 Davies-Bouldin Indices (DB Index) of clustering results by different weight combinations in defining a new distance between trajectories

	Combination 1	Combination 2	Combination 3
DB Index	4.88	7.52	7.67

세 가지 가중치 조합 모두에서 선박의 대지침로, 그리고 대지속력 분포의 차이를 함께 감안하면 군집별로 차이가 확인된다. 다만, 가중치 조합 1에서는 항적 대지속력뿐 아니라 대지침로의 분포에서도 군집별 차이가 명확하고, 군집 내 각 수치의 분산이 다른 조합에 비해 작다는 점이 확인된다. 또한 Table 1에서 확인할 수 있듯이, Davies-Bouldin 지표를 활용한 정량적 군집화 평가 척도에서도, 가중치 조합 1의 지표는 4.88로, 다른 조합의 7.5 또는 7.7에 비해 값이 낮다는 것을 알 수 있다. 이러한 결과를 종합하여, 본 연구에서 제안한 새로운 항적 간 거리 척도를 최적으로 정의하는 조합으로서 가중치 조합 1을 선택하였다.

5.2 기존 거리 척도와 새로운 척도 간 비교

항만에 접근하는 실제 선박 항적을 활용한 선행된 실험을 통해 새로운 항적 간 거리 척도를 정의하기 위한 최적의 가중치 조합을 확인하였다. 마찬가지로 방식으로 기존 항적 간 거리 측정 방식 대비 본 논문에서 제안하는 새로운 척도의 유효성을 실증하고자 2022년 11월 간 제주항에서 출항하여 목포항에 입항한 비어선 선박들의 242개의 항적들을 대상으로 세 가지의 거리 측정 방식으로 계산한 거리를 기반으로 군집화를 수행하였다. 즉, 하우스도르프 거리, 다이내믹 타임 워핑 거리, 그리고 식 (6)와 같이 계산되는 새로운 거리를 적용하여 각각의 거리 매트릭스를 계산하고, 동일한 군집화 알고리즘을 적용하는 과정을 다양한 관점에서 비교 분석하였다.

$$D_{new} = 0.5 * hausdorff(A, B) + 0.25 * (| SOG_{mean}^A - SOG_{mean}^B |) + 0.25 * (| COG_{var}^A - COG_{var}^B |) \quad (6)$$

우선 Table 2에서는 항적 간 거리 척도별 거리 매트릭스를 계산하는 데 소요된 연산 시간을 확인할 수 있다. 동일한 항적 데이터에 대해 하우스도르프 거리, 그리고 새로운 거리 척도에서는 거리 매트릭스 연산이 약 249초 정도가 소요되나, 다이내믹 타임 워핑 거리에 대해서는

그보다 2분 이상이 더 소요된 약 374 초가 기록되었다. 이는 비교적 연산량이 많다는 다이내믹 타임 워핑 거리에 대한 단점을 지적한 기존 연구들과 일치하는 결과이다(Magdy et al., 2015).

Table 2 Time consumed for distance matrices computation by different distance metrics (Unit: seconds)

	Hausdorff	Dynamic Time Warping	New Distance
Time	248.7	373.7	249.6

Fig. 3의 a), b), 그리고 c)는 각각 가중치 하우스도르프 거리, 다이내믹 타임 워핑 거리, 그리고 새로운 거리를 적용한 군집화의 결과이다. Fig. 1과 마찬가지로, 동일한 군집은 동일한 색의 웨이포인트들로 지도상에 표시되었다. Fig. 3의 a)에서 나타났듯이, 하우스도르프 거리를 활용한 군집화 결과는 거의 모든 항적들을 같은 군집으로 판단한 것을 알 수 있다. 한편 다이내믹 타임 워핑을 적용한 b)는 a)에 비해 좀 더 정교한 군집화 결과를 보여주고 있다. 특히 해당 결과에서는 비교적 해안선에 근접한 항적들은 여타 다른 항적과는 다른 군집으로 계산된 것을 알 수 있다. 한편 c)는 앞서 확인한 a)와 b)의 군집화 결과와 비교하여 특히 먼 바다에서 입항하는 항적들에 대해 좀 더 다양하게 군집화한 결과를 보여주고 있다.

5.1.과 마찬가지로 각 거리 척도에 기반한 항적 군집화 결과가 선박별 거동 정보를 반영하였는지 확인하기 위하여, Fig. 4과 같이 군집별 항적들의 평균 대지침로값과 평균 대지속력의 분포를 확인하였다. Fig. 4의 a)는 하우스도르프 거리 기반 군집화 결과에서는 대지침로 및 대지속력에서 군집별 차이가 거의 나타나지 않는다. 한편 다이내믹 타임 워핑 거리 기반 군집화 결과인 b)와 신규 거리 척도 기반 결과인 c)에서는 a)의 분포에 비해 평균 대지침로 및 대지속도에서 군집 간의 차이가 나타나고 있다.

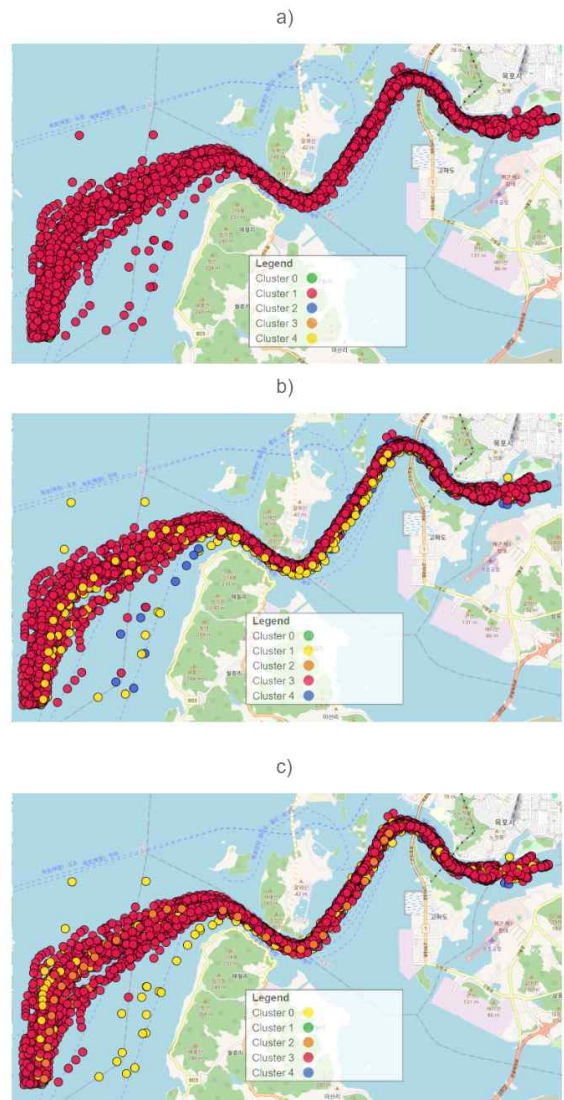


Fig. 3 Clustering results for different distance metrics between trajectories. a) Hausdorff distance b) Dynamic time warping distance c) New distance metric from the present article

거리 척도별 군집 간 거리와 군집 내 변동성을 정량적으로 확인하기 위하여 Table 3과 같이 Davies-Bouldin 지표를 활용하였다. 하우스도르프 거리에 기반한 군집화 결과의 지표는 약 20으로서 5.8인 다이내믹 타임 워핑, 그리고 6.1인 새로운 거리 척도에 의한 군집화 결과 각각에 비해 매우 높다. 이는 Fig. 4에서 시각화한 거리 척도별 군집의 동적 특성별 분포에서, 하우스도르프 거리 기반 군집에 비해 다이내믹 타임 워핑 거리 및 신규 거리 척도 기반 군집에서 특히 군집 간

연산 상의 효율성을 확보하였다는 점에서 실제적 가치가 높다고 할 수 있다.

Table 3 Davies-Bouldin Indices (DB Index) of clustering results by different distance metrics

	Hausdorff	Dynamic Time Warping	New Distance
DB Index	20.34	5.80	6.08

6. 결 론

본 연구는 항적 및 선박 특성을 유효하게 반영한 항적 간 거리 척도와 이를 기반으로 한 항적 군집화 방법론이 국내 연안 해역 환경에서의 해양 안전 서비스 발전에 기여할 수 있다는 점에 착안하여, 이에 부응할 새로운 거리 척도를 제안하였다. 이어서 신규 척도의 실제 가치를 검증하기 위하여 서남해안 항만을 출입한 실제 AIS 항적 데이터와 병합 군집화 알고리즘을 활용하여 항적 군집화 결과를 시각화하고 정량적으로 평가하는 등 여타 기존 항적 간 거리 척도와의 비교 분석을 다각적으로 수행하였다. 본 연구에서 제안한 신규 척도는 전통적으로 위치 시계열 간의 지리적 유사도를 측정하는 데 활용되는 하우스도르프 거리와 두 항적 간의 대지속력의 평균 간의 차이, 그리고 대지침로의 분산 간의 차이를 가중합하여 설계되었다. 실제 항적 데이터를 활용한 비교 분석 결과, 개발한 새로운 거리 척도를 활용한 항적 군집화 결과가 기존 척도에 비해 항적 간 지리적 위치의 차이나 대지속도 및 대지침로 등 선박 거동 특성의 분포 차이를 비슷하거나 그 이상의 수준으로 정교하게 반영하고 있음을 데이터 시각화로써 확인하였다. 정량적으로는 군집 간 또는 군집 내 변동을 기준으로 군집화의 질을 평가하는 Davies-Bouldin 지표를 기준으로, 더욱 우수하거나 비슷한 수준을 기록한 한편, 계산 효율성에서는 특히 우수함을 보였다.

본 연구에서 고안한 항적 간 거리 척도는 항적 간 지리적 유사도뿐 아니라 항적의 동적 특성을 포괄적으로 반영하였다. 이 척도를 기반으로 항적 군집화를 적용한다면, 다량의 복잡한 항적 데이터를 항적의 지리적 위치뿐 아니라 선박의 동적 특성까지 고려한 정교한 항적 군집들로 요약할 수 있으며, 각 항적 군집 내에서 다른 항적들과의 거리가 최소인 항적을 군집 대표 항적으로 추출할 수 있을 것이다. 다수의 선박이 몰려 혼잡한 연안 해역 해상교통관제 서비스에 부담을 주는 상황에서, 선박과 현재 항적의 동적 특성별로 가장 유사한 대표 항적을 신속히 제공한다면, 해상 안전 관리 업무의 부하를 경감할 뿐만 아니라 데이터에 의해 해상 안전이 보장된 항적으로 선박들을 유도함으로써 해양안전에 기여할 수 있을 것이다.

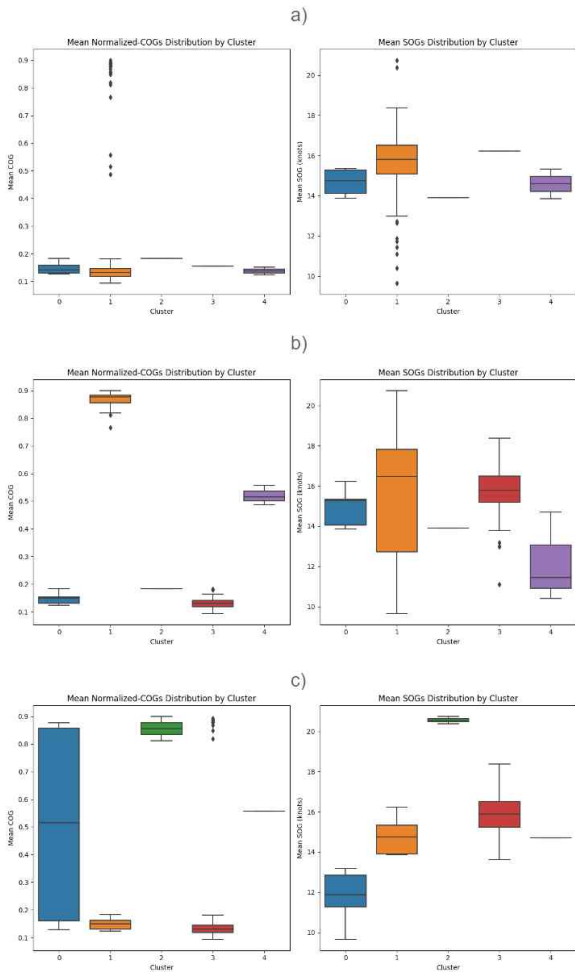


Fig. 4 Distributions of mean normalized COG values and mean SOG values in clustering results for different distance metrics between trajectories. a) Hausdorff distance b) Dynamic time warping distance c) New distance metric from the present article

차이가 두드러진다는 분석과 일관되는 결과이다.

항적 간 거리 척도의 실제적인 유용성은 그 가치가 다양한 측면에서 종합적으로 평가되어야 할 것이다. 군집의 질을 정량적으로 나타내는 Davies-Bouldin 지표에서는 다이내믹 타임 워핑 거리 기반의 군집화 결과가 새로운 거리 척도보다 근소하게 우수하다. 그러나 매트릭스를 계산하는 데 소요된 연산 시간에서는 새로운 항적 간 거리 척도가 2분 이상 효율적임을 확인하였다. 실제 선박 항적 데이터가 대용량 시계열 데이터임을 감안한다면, 그리고 데이터가 누적됨에 따라 서로 간의 거리를 계산해야 할 항적들이 늘어나게 됨을 감안한다면 신규 거리 척도의 계산 효율성은 더욱 가치가 있다. 따라서 본 연구에서 개발한 새로운 항적 간 거리 측정 방법론은 연산량이 많은 다이내믹 타임 워핑 거리 기반의 방법론과 비등한 수준의 항적 군집화 결과를 도출하면서도,

새로운 항적 간 거리를 실증하기 위하여 본 논문에서는 목포, 광양 등 두 가지 항만을 왕래한 AIS 항적 데이터만을 활용하였으나, 향후 연구에서는 보다 다양한 실험 조건들을 설계하여 국내 해역에서의 보편적인 유효성을 입증해야 할 것이다. 아울러 선박 위치, 대지속력 및 대지속도 정보 외에도 기상, 조위 등 해양 환경 변수를 추가로 고려한 항적 간 군집화 결과를 분석한다면 개발된 신규 거리 척도가 더욱 적시적인 해상 안전 서비스 구현에 기여할 수 있을 것이다.

후 기

본 논문은 해양수산부 재원으로 선박해양플랜트연구소의 주요사업인 "선박해양 디지털 전환 지원을 위한 디지털서비스 플랫폼 개발"에 의해 수행되었습니다(1525014879).

References

- [1] Alt, H.(2009), "The computational geometry of comparing shapes", *Efficient Algorithms: Essays Dedicated to Kurt Mehlhorn on the Occasion of His 60th Birthday*, pp. 235-248.
- [2] Buchin, M., Dodge, S. and Speckmann, B.(2012), "Context-aware similarity of trajectories", *Proc. of the Seventh International Conference on Geographic Information Science, GIScience*, pp. 43-56.
- [3] Davies, D. L. and Bouldin, D. W.(1979), "A cluster separation measure", *IEEE transactions on pattern analysis and machine intelligence*, Vol. PAMI-1, No. 2, pp. 224-227.
- [4] Dobrkovic, A., Iacob, M. E. and van Hillegersberg, J.(2018), "Maritime pattern extraction and route reconstruction from incomplete AIS data", *International Journal of Data science and Analytics*, Vol. 5, pp. 111-136.
- [5] Fisher, D.(1996), "Iterative optimization and simplification of hierarchical clusterings", *Journal of artificial intelligence research*, Vol. 4, pp. 147-178.
- [6] Keogh, E. and Ratanamahatana, C. A.(2005), "Exact indexing of dynamic time warping", *Knowledge and information systems*, Vol. 7, pp. 358-386.
- [7] Kim, K. I., Jeong, J. S. and Park, G. K.(2014), "A Study on Near-miss Incidents from Maritime Traffic Flow by Clustering Vessel Positions", *Journal of Korean Institute of Intelligent Systems*, Vol. 24, No. 6, pp. 603-608.
- [8] Lee, H. K., Chang, S. R., Jeong, G. N. and Park, Y. S.(2010), "A proposal on the marine traffic supporting system in VTS area", *Journal of Navigation and Port Research*, Vol. 34, No. 9, pp. 693-698.
- [9] Magdy, N., Sakr, M. A., Mostafa, T. and El-Bahnasy, K.(2015), "Review on trajectory similarity measures", *Proc. of the Seventh International Conference on Intelligent Computing and Information Systems, ICICIS*, pp. 613-619.
- [10] Oh, J. Y., Kim, H. J. and Park, S. K.(2018), "Detection of ship movement anomaly using AIS data: a study", *Journal of Navigation and Port Research*, Vol. 42, No. 4, pp. 277-282.
- [11] Pallotta, G., Vespe, M. and Bryan, K.(2013), "Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction", *Entropy*, Vol. 15, No. 6, pp. 2218-2245.
- [12] Park, S. Y. and Park, Y. S.(2022), "A basic study on intersection congestion using simulation: focusing on the crossing vessels in Busan New Port caution area", *Journal of Korean Society of Transportation*, Vol. 40, No. 3, pp. 319-334.
- [13] Rong, H., Teixeira, A. P. and Soares, C. G.(2020), "Data mining approach to shipping route characterization and anomaly detection based on AIS data", *Ocean Engineering*, Vol. 198, 106936.
- [14] Zhao, L. and Shi, G.(2019), "A novel similarity measure for clustering vessel trajectories based on dynamic time warping", *The Journal of Navigation*, Vol. 72, No. 2, pp. 290-306.
- [15] Zhen, R., Jin, Y., Hu, Q., Shao, Z. and Nikitakos, N.(2017), "Maritime anomaly detection within coastal waters based on vessel trajectory clustering and Naïve Bayes Classifier", *The Journal of Navigation*, Vol. 70, No. 3, pp. 648-670.

Received 30 November 2023

Revised 08 December 2023

Accepted 18 December 2023