

K-Means 클러스터링을 활용한 선박입항패턴 단계화 연구

이정석* · 이형탁** · † 조익순

*한국해양대학교 석사후 연구원, **한국해양대학교 해양과학기술전문대학원 박사과정생, † 한국해양대학교 해사글로벌학부 교수

A Study on Phase of Arrival Pattern using K-means Clustering Analysis

Jeong-Seok Lee* · Hyeong-Tak Lee** · † Ik-Soon Cho

*Researcher, Graduate School of Korea Maritime and Ocean University, Busan 49112, Korea

**Student, Ocean Science and Technology School, Korea Maritime and Ocean University, Busan 49112, Korea

† Professor, Division of Global Maritime Studies, Korea Maritime and Ocean University, Busan 49112, Korea

요 약 : 4차 산업혁명으로 인공지능, 사물인터넷, 빅데이터 등의 기술이 조선 해운 산업에 매우 밀접하게 연관 되고 있고 이는 자율운항선박의 탄생을 가져왔다. 현재 선박의 기술적 특성상 속력을 갑자기 낮출 수 없으므로 항만에 접안하기 위해 예인선의 도움, 도선사의 승선, 육상 관제센터의 선박 컨트롤 등 복잡한 커뮤니케이션을 필요로 한다. 본 연구에서는 자율운항선박이 도입될 경우 선박이 입항하기 위한 컨트롤 기준을 어떻게 설정할지 해결하고자 클러스터링 분석을 사용하였다. 입항 선박의 축적된 AIS 데이터를 기반으로 입항 패턴을 정량적으로 단계화하고자 K-Means 클러스터링을 사용했고 SOG(Speed over Ground), COG(Course over Ground), ROT(Rate of Turn)를 사용하여 입항 단계를 6개로 구분하였다.

핵심용어 : 자율운항선박, K-Means, 클러스터링, 입항패턴, 단계화, 선박자동식별시스템

Abstract : In 4th Industrial Revolution, technologies such as artificial intelligence, Internet of Things, and Big data are closely related to the maritime industry, which led to the birth of autonomous vessels. Due to the technical characteristics of the current vessel, the speed cannot be suddenly lowered, so complex communication such as the help of a tug boat, boarding of a pilot, and control of the vessel at the onshore control center is required to berth at the port. In this study, clustering analysis was used to resolve how to establish control criteria for vessels to enter port when autonomous vessels are operating. K-Means clustering was used to quantitatively stage the arrival pattern based on the accumulated AIS(Automatic Identification System) data of the incoming vessel, and the arrival phase using SOG(Speed over Ground), COG(Course over Ground), and ROT(Rate of Turn) Was divided into six phase.

Key words : Maritime Autonomous Surface Ship, K-Means, Clustering, Arrival Pattern, Phase, Automatic Identification System

1. 서 론

최근 4차산업의 혁명으로 인공지능(AI), 사물인터넷(IoT), 빅데이터(Big data)등의 기술이 조선해운 산업에 융화되고 자율운항선박의 탄생을 가져왔다. 현재 운항중인 선박들은 기술적 특성상 입항하기 위해 속력을 서서히 낮추거나 예인선의 도움을 받는다. 또한 도선사의 승선, 육상관제센터의 통항 선박의 컨트롤 등 복잡한 커뮤니케이션을 필요로 한다.

2. 분석 데이터

클러스터링 분석을 위해 Christopher et al.(2018)은 1년간 수집한 AIS 데이터를 기반으로 연구를 수행하였으며, 데이터의 용량은 1TB, 약 300억개로 구성되어 있다. 사용된 변수는 SOG(Speed over Ground)와 ROT(Rate of Turn) 이다. 본 연

구에서도 마찬가지로 AIS 데이터를 사용하였고 3일 동안 부산에서 수집하였다. 데이터 용량은 48MB, 약 100만개로 구성되어 있으며 사용된 변수는 SOG, ROT에 더불어 COG(Course over Ground)를 사용하였다. AIS 데이터는 해상 교통 및 혼잡에 대한 통계, 운영 관리, 다양한 항구로의 접근 분석, 해상 운송에 대한 지속 가능한 솔루션, 최적 항로, 특히 선박 항행 패턴을 얻는 방법으로 사용된다.

3. K-Means 클러스터링

인공지능의 한 분야인 기계학습은 지도학습, 비지도학습으로 크게 2개로 구분할 수 있다. 지도학습은 목적함수가 있고, 훈련을 통해 예측을 목적으로 한다. 인공신경망, 의사결정나무, 분류 분석, 앙상블 등이 해당된다. 한편, 비지도 학습은 목적함수가 없고, 데이터가 어떻게 구성되어 있는지 알아내는데 목적이 있

† 교신저자 : 종신회원, Ischo@kmou.ac.kr

* 종신회원, shoesy548@kmou.ac.kr

다. 클러스터링, SOM(Self Organization Map), 연관분석 등이 해당된다. 본 연구에서 사용된 분석은 클러스터링 분석이다. 또한, 클러스터링 분석은 계층적, 비계층적으로 나뉘며 K-Means 분석은 프로토타입의 미리 K개로 군집을 나누고 최적의 개수를 찾아가는 비계층적 분석방법으로 나타난다.

3.1 거리 산출

식 (1)과 같이 K-Means 클러스터링을 사용하여 어떻게 군집을 나누는지에 대한 기준은 거리 산출에 있다.

$$d_E(x,y) = \left[\sum_{j=1}^m |x_j - y_j|^2 \right]^{\frac{1}{2a}} \quad (1)$$

산포되어 있는 데이터 끼리 다양한 거리 산출 방법이 있으나 본 연구에서는 유클리드 거리를 사용하여 군집을 나타낸다. 또한, 최적의 군집수를 선정하기 위해 F통계량을 사용한 R^2 값으로 최적의 군집수를 Table 1과 같이 나타냈다.

Table 1 Characteristics of Variable

Variable	Mean	SD	R^2
SOG	4.29	4.08	0.97
COG	190.88	135.10	0.95

F통계량 기준의 R^2 값은 SOG의 경우 0.97, COG의 경우 0.95로 계산되었으며 1에 가까울수록 높은 정확성을 나타낸다. 즉, 해당 분석을 통한 최적 군집 개수는 22개로 나타났으며 입항 단계를 22단계로 구분은 한계점이 존재한다. 따라서 결측값의 삭제, 항해중인 선박, 부산항 내부의 특정 부두 대상, 그리고 군집의 개수를 Christopher et al.(2018)과 같이 6개로 선정하여 분석하였다. Fig. 1은 부산 신선대 부두를 기반으로 입항 패턴을 6개로 구분한 결과를 보여준다.



Fig. 1 Result of Arrival Pattern into Six Phase

4. 클러스터링 결과검증

본 연구는 6개의 클러스터링 분석 결과를 기반으로 SOG와 COG를 입항패턴 Matrix를 나타냈다. 데이터의 분포를 알아보기 위해 사분위 그래프를 사용하였으며 Fig. 2와 같이 결과가 산출되었다.

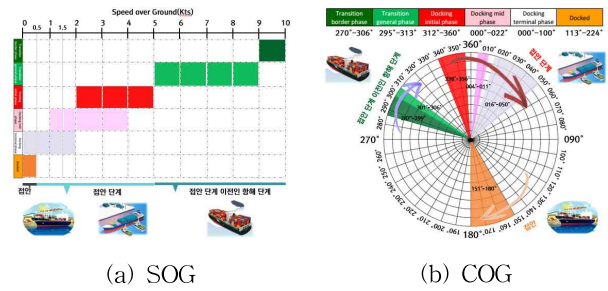


Fig. 2 Arrival Pattern Matrix based SOG and COG

5. 결 론

부산 북항 신선대 부두를 대상으로 K-Means 클러스터링 개수를 6개로 분석하였다. 사용된 데이터는 SOG, COG, ROT였으며 접안 단계에 따른 입항 패턴을 구분할 수 있었다. 속도를 줄여가며 선박이 접안하는 것은 당연한 이치이나, 데이터 기반과 정량적인 분석으로 자율운항선박의 입항 가이드라인에 적용할 수 있다는 점이 있다. 육상관제관점에서 접안 부두와의 거리를 산출하여 입항선박의 SOG와 COG의 권고가 가능할 것으로 판단된다. 향후 과제로 K-Means 클러스터링 분석 이외에도 다양한 클러스터링 분석을 수행하고자 하며, 대용량 데이터를 기반으로 분석해보고자 한다. 또한, 6개 클러스터링 이외에 국내 항만 실정에 알맞은 최적의 개수를 산출해보고자 한다.

사 사

본 논문은 2020년도 해양수산부 및 해양수산과학기술진흥원 연구비 지원으로 수행된 '자율운항선박 기술개발사업(20200615, 자율운항선박 육상제어 기술개발)'의 연구결과입니다.

참 고 문 헌

- [1] Christopher, B., Alex, N., Ioannis, T. and Sonia, W.(2018), "Analysing port and shipping operations using big data", June. 18.