

선박 대표 궤적 추출을 통한 Outlier 판별 알고리즘

박진관*, 오주성*, 김범무**, 전성민**, 이성로**, 정민아*

*목포대학교 컴퓨터공학과

**목포대학교 정보전자공학과

e-mail:chrispj@mokpo.ac.kr

Outlier Distinction Algorithm via Vessel Representative Trajectory Extraction

Jin-Gwan Park*, Joo-Seong Oh*, Bum-Mu Kim**, Sung-Min Jeon**,
Sung-Ro Lee**, Min-A Jeong*

*Dept of Computer Engineering, Mok-po University

**Dept of Information & Electronic Engineering, Mok-po University

요 약

본 논문은 해상 교통량 증가로 급증하는 선박 사고 위험을 줄이기 위해 안전 운항을 위한 대규모 선박 궤적 클러스터링을 제안한다. 선박의 위도와 경도, 이름 및 상태, 속도, 선수 방향 등이 기록된 대용량의 데이터집합을 바탕으로 선박 궤적 클러스터링을 통해 총 2개의 선박 대표 궤적을 추출한다. 해당 선박의 이전까지의 대표 궤적, 그리고 해당 해상의 모든 선박의 대표 궤적을 추출한 후 현재 해당 선박의 궤적패턴과 비교하여 유사하지 않으면 **Outlier**로 판별하여 이상 거동 및 불규칙 움직임, 충돌 상황을 대비할 수 있도록 의사결정에 도움을 줄 수 있는 알고리즘을 제안하였다.

1. 서론

최근 이동 단말기의 확산과 GPS(Global Positioning System), 위성, WSN(Wireless Sensor Network), RFID(Radio Frequency Identification) 등 통신 기술의 발달로 인하여 이동 객체의 궤적 정보 획득 수단이 다양해지고, 실제로 방대한 양의 위치 데이터가 발생하고 있다. 궤적 데이터는 이동 객체의 특정 시점에서의 위치를 연속적으로 기록한 데이터이다. 이런 궤적 데이터의 예로는 차량 이동 데이터, 선박 이동 데이터, 동물 추적 데이터, 태풍 경로 데이터 등 수없이 많다. 이러한 이동 객체의 궤적 데이터는 이동 객체의 행동을 분석 및 예측하기 위한 정보로 사용될 수 있다. 따라서 궤적 데이터 마이닝을 통한 궤적 데이터의 분석에 대한 관심이 높아졌으며, 여러 가지 이동 객체 궤적 데이터 마이닝 기법이 연구되었다[1][2].

선박 추적 시스템에서는 위성으로 측정된 사진이나 좌표 정보 및 AIS(Auto Identification System)를 기반으로 선박의 위치를 탐색 및 저장하고 이를 분류하는 기법을 통칭하며 어장 관리, 해양 오염 경로 예측, 밀입국 및 밀수 등을 관리하기 위한 국경 관리시스템 등에 사용된다. 이러한 시스템은 해로를 따라 이동하는 선박의 궤적 정보를 수집하여 일반적인 이동 경로를 분석하고, 이동의 목적 및 여러 조건에 따라 궤적 데이터를 분류하여 적용한다. 예를 들어 국경 관리 시스템의 경우, 관측된 선박의 경로가 궤적 데이터 마이닝을 통해 분석한 기존의 궤적 패턴과 유사하지 않은 경우, 이를 Outlier로 분류하고 면밀히 검토

하여 밀입국이나 테러 조직 등의 불법 이동 경로인지 확인할 수 있다[3]. 또한 궤적 데이터 마이닝을 통해 해양 사고가 다수 발생하는 지점 등을 미리 분석하여 선박 조난 시 구조대를 파급할 지역을 선정하는 등의 의사 결정에도 큰 도움을 줄 수 있다[1].

본 논문에서는 총 2개의 선박 대표 궤적을 추출하여 현재 해당선박의 궤적 패턴과 비교하는 방식을 제안한다. 2가지 선박 대표 궤적은 해당 선박의 이전까지의 대표 궤적과 현재 선박이 위치하고 있는 해상을 운항했던 모든 선박의 대표 궤적이다. 이 2가지의 궤적과 현재 선박의 궤적을 비교하여 현재 선박의 Outlier를 판별한 후 선박의 이상 거동 및 불규칙 움직임, 충돌 상황을 대비할 수 있도록 의사결정에 도움을 줄 수 있는 알고리즘이다[4]. 또한, 대규모의 AIS 데이터집합을 분석하기 위해 클라우드 컴퓨팅을 지원하는 아마존의 ec2에 분산처리 환경인 하둡 Map-Reduce를 적용하여 대용량의 데이터집합에 유연하게 대응하도록 하였다.

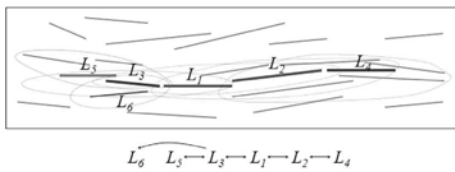
2. 관련연구

2.1 궤적 집단화(trajjectory clustering) 알고리즘

궤적 집단화 알고리즘은 밀도 기반 집단화 알고리즘에 기반한다. 밀도 기반 알고리즘에서 집단(cluster)은 밀도가 낮은 지역에 의해 분리된 밀도가 높은 지역으로 정의된다. 여기에서 밀도는 특정 반경 내에 있는 데이터 객체의 개수를 의미한다. DBSCAN을 비롯한 밀도 기반 알고리즘은

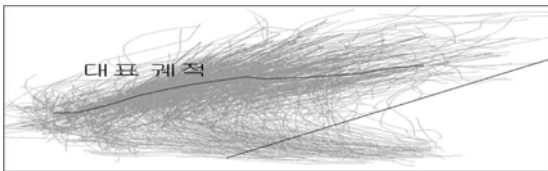
점 데이터를 위해 개발되었으며, 이들 알고리즘에서는 밀도를 계산할 때 점들 간의 거리만을 고려한다. 선분을 위한 밀도 기반 알고리즘에서는 선분들 간의 거리뿐만 아니라 선분의 모양까지도 함께 고려해야 한다. 즉, 거리가 가깝고 모양이 비슷한 선분이 있으면 그 선분을 현재의 집단에 계속 포함시켜 나간다. 이 과정을 더 이상 어떤 선분도 현재의 집단에 포함될 수 없을 때까지 반복한다.

그림 1은 밀도 기반 선분 집단화의 예를 나타낸다. 선분 L1에서 시작해 선분 L2와 선분 L3가 현재 집단에 포함되고, 선분 L2와 선분 L3의 이웃에 있는 선분들까지 현재 집단에 추가된다. 유사한 과정을 반복한 결과 선분 L1~L6의 타원에 포함되는 선분들이 하나의 집단으로 묶이게 된다. 그림 1에서 볼 수 있듯이 밀도 기반 집단화의 장점은 임의의 모양의 집단을 발견할 수 있다는 점이다.



(그림 1) 밀도 기반 선분 집단화의 예시

선분의 집단을 구한 후에, 각 집단 별로 그 집단을 대표하는 이동 형태를 구한다. 이를 대표 궤적(representative trajectory)라고 부른다. 그림 2에서 밀집된 선분들의 대표 궤적은 굵은 선으로 표시되어 있다. 대표 궤적이 바로 공통 부분 궤적을 나타낸다.



(그림 2) 대표 궤적의 예시

해당 기법은 지역적으로 유사한 궤적을 우선적으로 클러스터링 하고, 클러스터 내에서 가장 대표성을 띄는 궤적을 추출하기 때문에, 클러스터를 기반으로 유사도가 높은 패턴을 추출할 수 있는 장점을 지닌다[5]. 하지만, 모든 궤적들에 대하여 비교 연산을 수행하기 때문에, 상대적으로 많은 연산 비용을 요구한다. 그리고 모든 선박들의 궤적을 토대로 대표 궤적을 추출하기 때문에 해당 선박의 평소 항로와는 궤적 패턴이 상이할 수 있기에 Outlier를 판별하는 것이 쉽지 않다.

3. 대규모 선박 궤적 클러스터링

3.1 ec2 및 하둡 환경 설정 방법

Amazon EC2를 사용하여 하둡을 이용한 병렬 컴퓨팅 환경을 구축한다. Instance는 약 60개 정도를 사용하여 대용량의 데이터집합을 유연하게 처리하도록 한다. 각 Instance의 사양은 m3.xlarge (13 ECUs, 4 vCPUs, 2.5

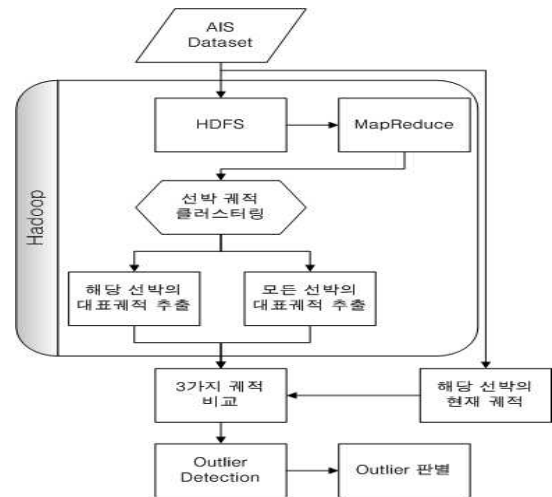
GHz, Intel Xeon E5-2670v2, 15 GiB memory, 2 x 40 GiB Storage Capacity)이다. Hadoop 1.2.1 버전을 통해 각 Instance를 노드로 설정하여 분산환경을 구축한다.

3.2 Dataset 특징 추출

대용량의 선박 위치 데이터집합은 <http://www.marinetraffic.com>의 Historical AIS data를 이용한다. Historical AIS data는 선박의 LON, LAT, VESSEL MMSI, STATUS, SPEED, COURSE, HEADING and TIMESTAMP (UTC) 등 상세한 정보가 기록 되어 있다. 본 연구에서는 2014년 1월부터 8월 까지 한국 해안의 선박 traffic을 수집한 AIS data set을 사용하여 선박 궤적을 분석하고자 한다.

3.3 Outlier 판별 알고리즘

본 논문에서는 모든 선박의 대표 궤적 및 해당 선박의 대표 궤적을 따로 추출하여 현재 궤적과의 비교를 통해 Outlier를 판별하는 알고리즘을 제안한다.

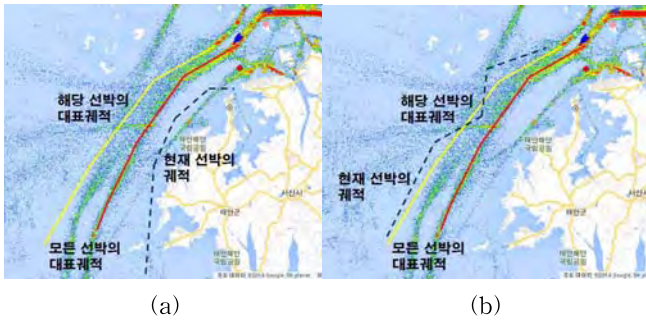


(그림 3) 제안 알고리즘

그림 3은 하둡과 맵 리듀스를 이용하여 선박 대표 궤적을 추출하는 알고리즘이다. 선박의 대표 궤적을 추출하기 위해 Historical AIS data의 데이터집합을 ec2에 구축해놓은 하둡의 HDFS에 저장한다. 그리고 궤적 집단화 알고리즘을 적용한 맵리듀스를 실행한다. 맵과 리듀스는 총 2번 반복하여 궤적을 추출한다. 첫 번째는 해당 선박의 과거 궤적을 바탕으로 대표궤적을 추출하고, 두 번째는 모든 선박의 과거 궤적을 바탕으로 대표궤적을 각각 추출한다. 그 후 추출된 대표궤적들과 해당 선박의 현재 궤적을 바탕으로 Outlier Detection을 수행한다. Outlier 판단은 해당 선박의 현재 궤적이 과거의 궤적 및 모든 선박의 대표 궤적 패턴과 유사하지 않으면 Outlier로 분류한다.

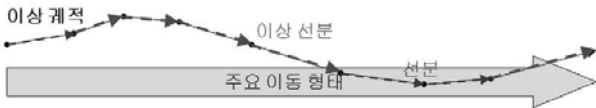
그림 4는 해당 선박의 현재 궤적, 해당 선박의 대표 궤적, 모든 선박의 대표 궤적을 가상으로 시뮬레이션한 결과를 나타낸 것이다. 그림 4의 (a) 상황처럼 해당 선박의 현재 궤적이 대표 궤적 및 모든 선박의 대표궤적과도 유사

하지 않으면 Outlier로 판별한다. 하지만 그림 4의 (b) 상황처럼 모든 선박의 대표궤적과는 유사하지 않아도 해당 선박의 대표궤적과 유사하면 Outlier로 판별하지 않는다.



(그림 4) 선박 대표 궤적 클러스터링 가상 시뮬레이션

Outlier 판별은 Outlier Detection을 통해 의사결정을 한다. 그림 5는 이상 궤적을 발견하는 단계를 보여준다. 주요 이동 형태와 동 떨어져 있는 선분은 주요 이동 형태와 상이하기 때문에 이상 궤적으로 판별된다. 그리고, 이들 이상 선분이 포함되어 있는 해당 궤적은 이상치로 판별된다. 따라서, 전체 궤적이 특이하지 않고 일부분만 특이하다더라도 효과적으로 이상 궤적으로 발견될 수 있다[3].



(그림 5) Outlier Detection

본 논문에서는 Outlier Detection을 수행할 때 주요 이동 형태를 선박의 대표 궤적 하나만을 토대로 Outlier를 판별하지 않고, 추출한 2개의 궤적을 이용하여 Outlier Detection을 수행한다.

4. 결론

해상을 운항하는 선박의 위치를 측위하는 기술이 늘어남에 따라 축적되는 데이터가 대규모로 형성되고 있고 이러한 데이터를 토대로 안정운항을 위한 데이터마이닝 연구가 진행되고 있다. 본 논문에서는 대규모로 축적된 선박 위치 데이터집합을 이용하여 선박 궤적 클러스터링을 한 후, 2가지의 선박 대표 궤적을 추출하는 알고리즘을 제안하였다. 해당 선박의 이전까지의 궤적, 현재 선박이 위치하고 있는 해상을 운항했던 모든 선박의 궤적을 추출한 후 현재 해당 선박의 궤적패턴과 비교하여 유사하지 않으면 Outlier로 판별함으로써 선박의 이상 거동 및 불규칙 움직임, 충돌 상황을 대비할 수 있도록 의사결정에 도움을 줄 수 있는 알고리즘이다.

향후에는 제안한 추출된 선박의 대표 궤적이 선박 안전 운항에 얼마만큼 영향을 미치는지에 대한 타당성 평가를 거쳐야 한다.

ACKNOWLEDGMENT

이 논문은 미래창조과학부 및 정보통신산업진흥원의 IT융합 고급인력과정 지원사업(NIPA-2014-H0401-14-1009)과 2014년도 정부 (교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2009-0093828)의 연구 결과로 수행되었음.

참고문헌

- [1] 장미영, 윤민, 장재우, “위치 기반 서비스 환경에서 이동 객체 데이터 마이닝 기법에 대한 고찰”, ISSN 1598-9798. 데이터베이스연구. 28권 제1호 2012년 4월
- [2] Gerben Klaas Dirk de Vries,, Maarten van Someren, “Machine learning for vessel trajectories using compression, alignments and domain knowledge”, Expert Systems with Applications Volume 39, Issue 18, Pages 13426 - 13439, 15 December 2012
- [3] J. Lee, J.Han, X. Li, “Trajectory Outlier Detection: A Partition-and-Detect Framework”, International Conference on Data Engineering, pp140-149, 2008
- [4] 정중식, 박계각, 김은경, “AIS 데이터를 활용한 선박궤적의 분석”, 한국항해항만학회 학술발표대회자료 2012, 38-40 , 2012
- [5] 이재길, “대규모 궤적 데이터를 위한 데이터 마이닝 툴”, 한국정보과학회, 정보과학회논문지 : 컴퓨팅의 실제 및 레터 15(3), 2009.3, 145-153 (9 pages)