기상데이터를 통한
닻 끌림 예측 모델 개발

2023 날씨 빅데이터 콘테스트 - 대폭발

목차

공모 배경

닻 끌림 예측 모델 개발 배경

닻 끌림이란?

강한 바람이나 파도 등을 맞으면 선박이 정박지에 머무르지 못하고 밖으로 밀리게 되는 현상

문 제

선박교통관제사의 눈으로
닻 끌림 여부를 일일이 확인하여,
작업 시간 및 정확도 면에서
많은 어려움 발생

해 결

정박선의 움직임(항적)의
학습을 기반으로
'닻 끌림 자동탐지 시스템' 개발
중에 있으나 정확도 향상 필요

개선 방향

예측 및 탐지 정확도 향상을
위하여 기상 데이터를 더한
예측 모델 개발 필요

2013년 10월 15일

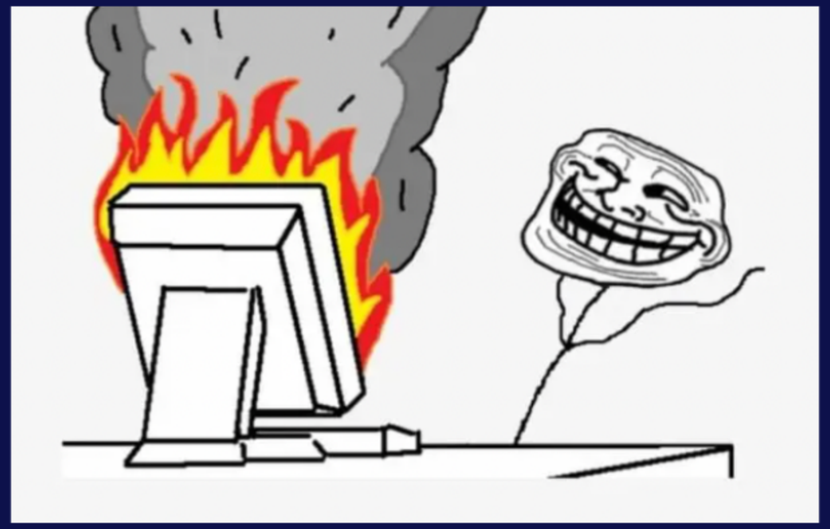
청루15호, 닻끌림으로 인한 방파제 충돌

9명 사망 2명 실종

벙커C유 106.7t 등 유출

동해면 입암1리 해안에 유막이 형성

임암리~발산리 해안가 4곳이 오염



VTS: 2022년 4분기 닻 끌림 탐지 알고리즘 구현 (0.7)

닻끌림 데이터

기상 데이터

닻끌림 데이터

해양 데이터

활용 데이터: 선박 데이터

test data

테이블 명	열 목록
울산_정박_닻끌림_test	ulsan_anch_drag_test.area
	ulsan_anch_drag_test.time
	ulsan_anch_drag_test.num
	ulsan_anch_drag_test.lat
	ulsan_anch_drag_test.lon
	ulsan_anch_drag_test.sog
	ulsan_anch_drag_test.cog
	ulsan_anch_drag_test.hdg

테이블 명	열 목록
부산_정박_닻끌림_test	busan_anch_drag_test.area
	busan_anch_drag_test.time
	busan_anch_drag_test.num
	busan_anch_drag_test.lat
	busan_anch_drag_test.lon
	busan_anch_drag_test.sog
	busan_anch_drag_test.cog
	busan_anch_drag_test.hdg

train data

테이블 명	열 목록	테이블 명	열 목록
울산_정박_train	ulsan_anch_train_final.num	부산_정박_train	busan_anch_train_final.num
	ulsan_anch_train_final.time		busan_anch_train_final.time
	ulsan_anch_train_final.latitude		busan_anch_train_final.latitude
	ulsan_anch_train_final.longitude		busan_anch_train_final.longitude
	ulsan_anch_train_final.sog		busan_anch_train_final.sog
	ulsan_anch_train_final.cog		busan_anch_train_final.cog
	ulsan_anch_train_final.hdg		busan_anch_train_final.hdg
울산_닻끌림_정답	ulsan_answer.area	부산_닻끌림_정답	busan_answer.area
	ulsan_answer.year		busan_answer.year
	ulsan_answer.num		busan_answer.num
	ulsan_answer.mon		busan_answer.mon
	ulsan_answer.day		busan_answer.day
	ulsan_answer.hour		busan_answer.hour
	ulsan_answer.min		busan_answer.min
	ulsan_answer.lat		busan_answer.lat
	ulsan_answer.lon		busan_answer.lon
울산_닻끌림_train	ulsan_drag_train_final.num	부산_닻끌림_train	busan_drag_train_final.num
	ulsan_drag_train_final.time		busan_drag_train_final.time
	ulsan_drag_train_final.latitude		busan_drag_train_final.latitude
	ulsan_drag_train_final.longitude		busan_drag_train_final.longitude
	ulsan_drag_train_final.sog		busan_drag_train_final.sog
	ulsan_drag_train_final.cog		busan_drag_train_final.cog
	ulsan_drag_train_final.hdg		busan_drag_train_final.hdg

활용 데이터: 해양/기상 데이터

해양데이터 (한국수력원자력)

test data

테이블 명	열 목록
KHNP_Buoy_test	khnp_buoy_test.yyyymmddhhmi
	khnp_buoy_test.stn_name
	khnp_buoy_test.ws
	khnp_buoy_test.wd

train data

테이블 명	열 목록
KHNP_Buoy_train	khnp_buoy_train.yyyymmddhhmi
	khnp_buoy_train.stn_name
	khnp_buoy_train.ws
	khnp_buoy_train.wd

해양데이터 (해양조사원)

test data

테이블 명	열 목록
KHOA_Buoy_test	khoa_buoy_test.yyyymmddhhmi
	khoa_buoy_test.stn_name
	khoa_buoy_test.ws
	khoa_buoy_test.wd_point
	khoa_buoy_test.wd

train data

테이블 명	열 목록
KHOA_Buoy_train	khoa_buoy_train.yyyymmddhhmi
	khoa_buoy_train.stn_name
	khoa_buoy_train.ws
	khoa_buoy_train.wd_point
	khoa_buoy_train.wd

기상데이터 (기상청)

test data

테이블 명	열 목록
KMA_PagoBuoy_test	kma_pagobuoy_test.yyyymmddhhmi
	kma_pagobuoy_test.stn
	kma_pagobuoy_test.stn_name
	kma_pagobuoy_test.max_wh
	kma_pagobuoy_test.sig_wh
	kma_pagobuoy_test.mean_wh

train data

테이블 명	열 목록
KMA_PagoBuoy_train	kma_pagobuoy_train.yyyymmddhhmi
	kma_pagobuoy_train.stn
	kma_pagobuoy_train.stn_name
	kma_pagobuoy_train.max_wh
	kma_pagobuoy_train.sig_wh
	kma_pagobuoy_train.mean_wh

데이터 전처리

정박 데이터 (울산_정박_train, 부산_정박_train) 제외

정박 데이터의 일부 열(sog, cog, hdg)의 경우, 닻 끌림 학습 데이터와 중복되었으며
현 위치에 데이터가 유의미한 영향을 미치지 않는다고 판단함

1. 경도와 위도 데이터의 특이성

2. 위치 데이터의 중복

데이터 전처리

정박 데이터 (울산_정박_train, 부산_정박_train) 제외

정박 데이터의 일부 열(sog, cog, hdg)의 경우, 닻 끌림 학습 데이터와 중복되었으며
현 위치에 데이터가 유의미한 영향을 미치지 않는다고 판단함

latitude & longitude 형식 변경 (예시: N000 → 000 / E000 → 000)

모델링 성능 개선 및 계산의 용이성을 위해 문자열 형태 변경 (character → numeric)

```
data_0$lat_N <- str_replace(data_0$lat_N,'N','')
data_0$long_E <- str_replace(data_0$long_E,'E','')
```

날짜 형식 통일 (yyyy-mm-dd hh:mm, col name: date)

각기 다른 날짜 형식을 가진 테이블의 데이터 통합을 위해 날짜 형식 통일

```
data_0$date <- sprintf("%04d-%02d-%02d %02d:%02d", data_0$year, data_0$month, data_0$day, data_0$hour, data_0$minute)
data_1$date <- sprintf("%04d-%02d-%02d %02d:%02d", data_1$year, data_1$month, data_1$day, data_1$hour, data_1$minute)
```

```
khoa$date <- strptime(khoa$date, format = "%Y%m%d%H%M")
khoa$date <- format(khoa$date, format = "%Y-%m-%d %H:%M")
```

데이터 전처리

기상 데이터 통합 (울산&부산_닻끌림_train&정답 + KHNP + KHOA + KMA)

통일된 날짜 형식으로 데이터 통합 (LEFT JOIN BY DATE)

```
# Perform left join
busan_final_merge <- left_join(main_khoa_khnp, kma_with_minutes, by = "date")
```

1. 울산 및 부산 닻끌림 데이터 LEFT JOIN KHOA BY DATE
2. 1번 LEFT JOIN KHNP BY DATE
3. 2번 LEFT JOIN KHOA BY DATE
4. 3번 LEFT JOIN KMA BY DATE

결측치 처리 (-99, 99.9, -999)

분석 결과 왜곡 방지 (분석 정확성 향상) 위해 결측치 제거

```
complete_rows <- complete_rows[!apply(complete_rows == -99, 1, any), ]
complete_rows <- complete_rows[!apply(complete_rows == -99.9, 1, any), ]
complete_rows <- complete_rows[!apply(complete_rows == -999, 1, any), ]
```

데이터 전처리

중복값 제거

분석 결과 왜곡 방지 (분석 정확성 향상) 위해 중복값 제거

```
complete_rows <- busan_final_merge[complete.cases(busan_final_merge), ]
complete_rows <- complete_rows[!duplicated(complete_rows), ]
```

분석 기법 및 결과

target: 닻 끌림 여부 (0: 닻 끌림 미발생 / 1: 닻 끌림 발생)

지도형 기계학습 (supervised learning) 중 이진 분류 (binary classification) 인 현 데이터 상황에 따라 아래와 같은 모델링 기법 사용

1. 결정나무 (decision tree)
2. 랜덤포레스트 (random forest)
3. XGBoost

평가 방법: ROC (Receiver Operating Characteristic Curve)

모델 선택

앙상블 학습 (Ensemble Learning)

1. Bagging
    a. Decision Tree: fast to train and predict
    b. Random Forest: excellent performance but slow to train
2. Boosting
    a. XGBoost: excellent performance but slow to train

분석 기법 및 결과

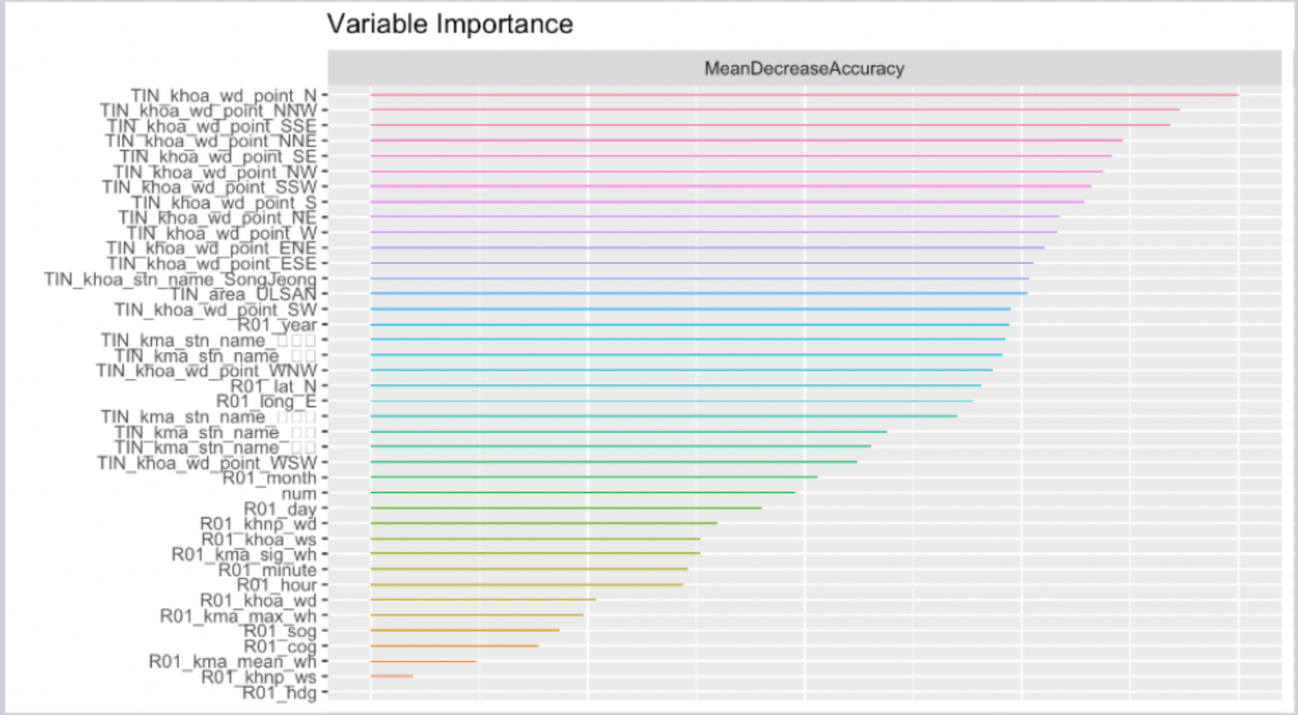
모델 별 성능 비교

Model	ROC
Decision Tree	0.6531
Random Forest	0.7921
XGBoost	0.7549

ROC 값이 가장 높은 Random Forest 모델링 기법 활용

분석 기법 및 결과

모델 해석



Variable Importance

	0	1	MeanDecreaseAccuracy	MeanDecreaseGini
R01_hdg	14.62	11.76	14.91	11.26
R01_khnp_ws	13.74	10.82	14.13	8.96
R01_kma_mean_wh	12.88	5.23	12.94	2.57
R01_cog	10.39	14.68	11.77	11.35
R01_sog	11.17	16.56	11.37	7.30
R01_kma_max_wh	10.86	6.68	10.91	3.48
R01_khoa_wd	10.49	4.37	10.68	7.63
R01_hour	8.83	8.89	9.06	3.15
R01_minute	7.47	10.63	8.97	7.68
R01_khoa_ws	8.59	2.84	8.73	8.52
R01_kma_sig_wh	8.68	7.08	8.73	3.05
R01_khnp_wd	7.78	12.87	8.40	8.01
R01_day	7.52	7.66	7.56	2.52
num	6.90	10.27	6.93	5.92
R01_month	6.47	9.38	6.52	1.86
TIN_khoa_wd_point_WSW	5.88	-2.99	5.78	0.67
TIN_kma_stn_name_당사	5.51	1.54	5.53	0.64
TIN_kma_stn_name_기장	5.28	-2.19	5.22	0.98
TIN_kma_stn_name_다대포	3.91	-0.74	3.92	0.36
R01_long_E	3.60	4.69	3.61	16.44
R01_lat_N	3.45	6.72	3.46	17.37
TIN_khoa_wd_point_WNW	3.23	0.00	3.23	0.56
TIN_kma_stn_name_장안	3.11	-3.70	3.06	0.31
TIN_kma_stn_name_오륙도	3.07	-2.01	3.00	0.33
R01_year	2.92	3.75	2.93	0.51
TIN_khoa_wd_point_SW	2.99	-2.00	2.89	0.34
[illegible]	2.50	0.47	2.59	0.51
TIN_khoa_stn_name_SongJeong	2.91	-5.09	2.55	0.68
TIN_khoa_wd_point_ESE	2.47	0.42	2.48	0.42
TIN_khoa_wd_point_ENE	2.53	-3.09	2.26	0.67
TIN_khoa_wd_point_W	1.64	4.31	2.04	0.45
TIN_khoa_wd_point_NE	1.94	-0.05	1.98	0.51
TIN_khoa_wd_point_S	1.51	0.00	1.51	0.03
TIN_khoa_wd_point_SSW	1.38	0.00	1.38	0.06
TIN_khoa_wd_point_NW	1.15	-0.21	1.17	0.42
TIN_khoa_wd_point_SE	1.00	0.00	1.00	0.27
TIN_khoa_wd_point_NNE	1.10	-3.20	0.79	0.39
TIN_khoa_wd_point_SSE	0.21	-1.73	-0.10	0.18
TIN_khoa_wd_point_NNW	-0.21	-1.00	-0.28	0.11
TIN_khoa_wd_point_N	-1.27	-0.68	-1.38	0.18

랜덤 포레스트의 주요 변수는 hdg, ws, wh, cog, sog 이며, 특히 hdg 및 파고가 주요 변수로 보여짐.

반면, wd point 및 station name 변수의 가중치는 저조함을 확인

활용 방안 및 기대 효과

활용 방안

사전에 해당 지역에 등록된
선박에 대한 비상연락망 체계 구축

→

기상 및 해양데이터를 활용한
‘닻 끌림 예측 모델’ 과 해상교통관제의
‘닻 끌림 자동탐지 시스템’을 활용하여
더 높은 정확도로 닻 끌림 발생을 예측

→

혼합 모델에서 30분 이상 지속적인
닻 끌림 예측 발생 시,
비상 연락망을 통하여 선장 및
주민들에게 **재난 문자를 전송**

↓

재난 문자 발송 후, 모든 선박에 대하여
항해 금지 및 관리 인력을 통한 항구통제

←

지속적인 닻 끌림 미예측
데이터 발생 1시간 후 통제 해제

←

통제 해제 후,
피해 상황 종합 및 사후 처리

활용 방안 및 기대 효과

기대 효과

1. 약 80~90% 확률에 가까운 닻 끌림 예측을 통해 해양사고 및 인명피해 예방 가능
2. 관측 방식의 변화(수기에 가까운 관측 방법 → 인공지능 및 데이터 분석)를 통해 1인당이 지점/지역에 대한 관리 범위 증가(관리 인력 감소)

한계 및 개선점

1. 닻 끌림 예측 실패를 보완하기 위한 추가적인 방안 필요
2. 통제 관리 인력 추가 필요(대기/당직)
3. 선박의 관리 및 통제에 대한 매뉴얼 필요