J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr

심층신경망을 통한 해파리 출현 예측☆

The prediction of appearance of jellyfish through Deep Neural Network

황 철 훈[1] 한 명 묵[1]*
CHEOLHUN HWANG Myung-Mook Han

요 약

본 논문은 지구온난화로 인하여 수온이 상승되며 증가한 해파리의 피해를 감소하고자 연구를 진행하였다. 해수욕장에서 해파리의 등장은 해파리의 쏘임 사고로 인한 인명피해와 폐장으로 인한 경제적 손실이 발생할 수 있다. 본 논문은 선행 연구들로부터 해파리의 출현 패턴을 머신러닝을 통하여 예측 가능함 확인하였다. SVM을 이용한 해파리 출현 예측 모델 연구를 확대하여 진행하였다. 심층신경망을 이용하여 해파리 출현 유무 예측인 이진 분류에서 지수화 된 방법인 다중 분류로 확장하고자 하였다. 수집된 데이터의 크기가 작다는 한계점으로 인하여 84.57%라는 예측 정확도의 한계를 부트스트래핑을 이용하여 데이터 확장을 통해 해결하고자 하였다. 확장된 데이터는 원본 데이터보다 약 7% 이상의 높은 성능을 보여주었으며, Transfer learning과 비교하여 약 6% 이상의 좋은 성능을 보여주었다. 최종적으로 테스트 데이터를 통하여 해파리 출현 예측 성능을 확인한 결과, 해파리의 출현 유무를 예측할 시 높은 정확도로 예측이 가능함을 확인하였으나, 지수화를 통한 예측에서는 의미 있는 결과를 얻지 못하였다.

☞ 주제어 : 심층신경망, 부트스트래핑, 데이터 확장, 전이학습, 랜덤포레스트, 연속 전진 선택법

ABSTRACT

This paper carried out a study to reduce damage from jellyfish whose population has increased due to global warming. The emergence of jellyfish on the beach could result in casualties from jellyfish stings and economic losses from closures. This paper confirmed from the preceding studies that the pattern of jellyfish's appearance is predictable through machine learning. This paper is an extension of The prediction model of emergence of Busan coastal jellyfish using SVM. In this paper, we used deep neural network to expand from the existing methods of predicting the existence of jellyfish to the classification by index. Due to the limitations of the small amount of data collected, the 84.57% prediction accuracy limit was sought to be resolved through data expansion using bootstraping. The expanded data showed about 7% higher performance than the original data, and about 6% better performance compared to the transfer learning. Finally, we used the test data to confirm the prediction performance of jellyfish appearance. As a result, although it has been confirmed that jellyfish emergence binary classification can be predicted with high accuracy, predictions through indexation have not produced meaningful results.

☞ keyword : Deep Neural Network, Bootstrapping, Data Expansion, Transfer learning, Random Forest, Sequence Forward Selection

1. 서 론

해파리의 집중 출현 기간과 우리나라의 여름 휴가철이 맞물리며 해수욕장의 해파리 쏘임 사고가 발생하여 국민 안전을 위협하고 있다. 만약 해파리의 출현으로 인하여 해운대 해수욕장이 1일 폐장을 하게 될 경우, 해운대 해수욕장은 111억 원 이상의 손실이 발생하여 경제적 피해로 확대되게 된다[1].

정부에서는 해파리 피해를 예방하는 관점에서 국립수산과학원을 통하여 해파리 모니터링 주간보고를 통한 모니터링을 강화 및 실태를 조사하고 있으며, R&D 등을 통하여 피해 방지대책 및 수거사업을 추진해왔다. 그 결과 표 1과 같이 해파리 포획 건수와 마리수가 줄어들었으며, 해파리 쏘임 피해도 줄어든 것을 확인할 수 있다. 그러나 2018년 기록적인 폭염으로 인하여 해수욕장의 관광객은 줄어든 상황에서 해파리 쏘임 사고가 작년에 비해 약 56%로 증가하였다.

본 논문은 해파리 출현 패턴을 통하여 예측할 수 있다는 선행 연구들을 토대로 공공기관으로부터 수집되고 있는 해양 데이터들을 활용하여 해파리 출현 예측이 가능함을 확인하고자 연구를 진행하며, 해수욕장 해파리 쏘임 사고 예방 및 기타 피해를 최소화 시키고자 한다.

1 Department of Software, Gachon University, Seongnam-si, 13120, Korea.
* Corresponding author (mmhan@gachon.ac.kr)
[Received 25 March 2019, Reviewed 25 April 2019(R2 9 June 2019), Accepted 19 July 2019]
☆ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2019년도 문화기술연구개발 지원사업으로 수행되었음

(표 1) 해파리 쏘임 및 포획
(Table 1) Jellyfish stings and capture

년 도	해파리 쏘임		해파리 포획	
	건	명	건	마리
2013	2,160	2,160	373	919
2014	41	41	47	2,212
2015	119	121	333	10,230
2016	360	361	1,128	2,186
2017	67	67	251	3,097
2018	105	105	278	709
합 계	2,852	2,855	2,410	19,353

본 구성은 다음과 같다. 제 2장과 3장에서 관련 연구 및 제안하는 예측 시스템에 대하여 설명한다. 제 4장을 통해 데이터 전처리 방법을 제시한다. 제 5장에서는 실험의 한계와 이를 해결하기 위한 데이터 확장 방법을 제시한다. 제 6장에서는 보완된 실험 및 결과를 기술하며, 마지막으로 7장에서 결론을 내린다.

2. 관련 연구

2.1 해파리 출현 예측

우리나라에서 진행된 선행 연구에서는 표층수온, 풍속, 우천일수를 해파리 출현과 관련된 특징 변수로 하여 연구를 진행하였으며[2], 일본에서는 수온과 염분을 이용하여 노무라입깃해파리의 출현 및 가능성을 예측하였다[3]. 이외의 총 6개의 연구를 고찰하여 수온, 풍속, 풍향, 우천일수, 염분, 동물성 플랑크톤, 수소 농도, 엽록소, 자외선, 용존산소 등의 특징들을 통하여 해파리 출현 패턴을 머신러닝을 이용하여 예측하고자 하였다.

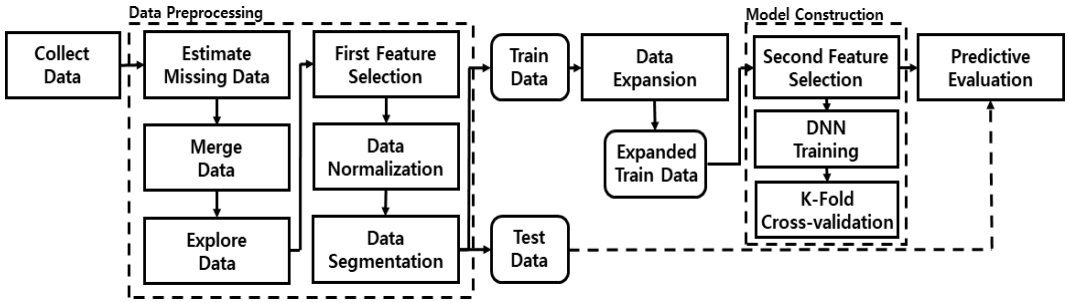
2.2 심층신경망

심층신경망(Deep Neural Network, DNN)은 신경망에 확장된 머신러닝 방법이다. 입/출력 계층 사이에서 여러 은닉 계층이 포함된 구조를 갖고 있어 여러 카테고리의 분류가 가능하다. 각 뉴런들로부터 산출된 값들은 활성 함수를 통하여 선형 또는 비선형 분류할 수 있어 이미지 분류, 음성인식 등 여러 분야에 활용되고 있다[4].

선행 연구를 통해 SVM을 이용하여 해파리 출현 유무에 대해 예측하였다. 본 연구에서는 출현 유무에 발전하여 지수화를 통한 예측 결과를 도출하는 것을 목표로 한다. 이에 SVM보다 심층신경망을 이용하는 것이 지수화 도출에 적절하다고 판단하여 활용하고자 하였다.

3. 제안 시스템

해파리 출현 예측 흐름은 그림 1과 같이 정의하였다. Collect Data 단계에서 수집된 데이터들은 데이터 전처리 단계인 Data Preprocessing 단계를 거친다. 전처리 단계에서는 결측 데이터를 보정하는 Estimate Data, 수집 기간을 통합하는 Merge Data, 특징 변수별 데이터 탐색 과정인 Explore Data, 불필요한 특징 변수를 삭제하는 Data Cleaning, 데이터를 정규화하는 Data Normalization, 훈련 데이터(Train Data)와 테스트 데이터(Test Data)를 나누는 Data Segmentation으로 순차적으로 진행한다. 훈련 데이터는 데이터 크기의 한계점을 극복하기 위해 부트스트래핑 이용한 데이터 확장 단계인 Data Expansion을 거쳐 Expanded Data를 만들게 된다. Model Construction 단계에서 Feature Selection을 통해 특징 변수를 선택하고 DNN Training에서 심층신경망을 통해 학습하게 된다. K-Fold Cross-validation에서 검증 과정 거쳐 높은 정확도의 예측



(그림 1) 해파리 출현 예측 흐름
(Figure 1) The prediction flow of jellyfish appearance

모델을 Predictive Evaluation 단계에서 테스트 데이터를 통해 예측 결과를 도출하고 평가한다.

4. 데이터 전처리

4.1 Estimate Missing Data

해양 수집 데이터는 일부 관측소에 따라 수집되지 않은 기간이 존재할 수 있다. 해파리 출현 데이터인 해파리 모니터링 주간보고의 단위는 주간 단위이기 때문에 데이터 통합 시 학습할 데이터의 크기가 적어진다. 수집된 데이터를 최대한 이용할 수 있도록 해양 및 기상 관측 데이터의 연속성 및 군집화의 특징을 이용하여 결측 데이터를 추정하는 방법[5]을 통해 데이터를 보완하였다.

4.2 Merge Data

수집된 해양 데이터는 분간 데이터이다. 이를 주간데이터로 통합하기 위해 기상청의 방재기상관측연보에 기술된 통계산출 방식을 적용하였다. 각 특징 변수 별 주간 평균, 주간 평균 최고, 주간 평균 최저, 주간 최고, 주간 최저의 값을 구하였다.

해파리 출현 정보는 해파리 출현 유무인 이진 데이터로 예측하던 방식을 지수화하여 예측하도록 하였다. 중국 황하 국제공항, 호주 아델레이드 공항당국, 국내 제주 공항의 조류위험평가모델를 참고하여 국립 수산과학부의 주간 해파리 모니터링 보고를 표 2를 기준으로 0부터 4까지 총 5가지의 지수로 구분하였다.

4.3 Explore Data & Cleaning

병합된 데이터에 대한 특성을 확인하기 위하여 탐색 과정을 진행하였다. 각 특징 변수들의 평균, 중앙값, 표준편차를 살펴보았으며 일부 특징 변수에서 평균, 중앙값, 표준편차가 모두 0인 경우를 확인할 수 있었다. 이는 주간 평균 최소와 주간 최소 특징 변수에서 나타났다. 해당 특징 변수는 해파리 출현 예측 간에 영향력이 없을 것으로 판단하여 삭제하였다.

(표 2) 해파리 출현 예측 지수
(Table 2) jellyfish appearance prediction index

구분	해파리 출현 지수				
해파리 주의보	-	미발령	미발령	미발령	발령
거리	-	근해	연안	연안	연안
밀도	미출현	저/고	저밀도	고밀도	고밀도
지수	0	1	2	3	4

4.4 Data Normalization & Segmentation

특징 변수인 기압 수치는 1000~1030 사이이며, 조위 수치는 60~80 사이, 풍속 수치는 3~7 사이 등으로 데이터 스케일 격차가 크다. 심층신경망을 통하여 해파리 출현을 예측할 경우 순전파, 역전파 시 가중치에 영향을 주어 좋은 결과를 내지 못한다. 이를 해결하기 위해 각 특징 변수별 Min-Max Normalization을 통하여 0에서 1사이로 데이터를 정규화를 하였다.

총 258개의 데이터 세트 중에서 2014년 1월부터 9월까지 총 247개의 행을 훈련 데이터, 그 이후 10월부터 11개의 행을 테스트 데이터로 구분하였다.

5. 실험 한계 및 데이터 확장

5.1 Experimental Limit

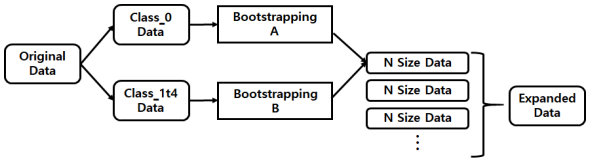
심층신경망을 통하여 해파리 출현 예측을 검증한 결과 최대 84.57%라는 한계를 가졌다. 첫 번째 원인은 심층신경망을 통하여 훈련을 할 만큼의 충분한 데이터 세트를 갖추지 못하였기 때문으로 보았다. 두 번째 원인은 노무라입깃해파리, 보름달물해파리, 유령해파리, 기타 해파리가 출현하는 데이터의 비율(해파리 출현 예측 지수 1 이상인 비율)이 전체 비율의 각각 15.38%, 24.63%, 8.86%, 4.87%로 낮았기 때문으로 보았다.

한계를 극복하기 위해서는 작은 사이즈의 데이터를 확장해야 하며, 해파리가 출현하는 데이터의 비율을 늘릴 필요가 있다. 이에 본 논문은 부트스트래핑을 통하여 해당 문제들을 해결하고자 하였다.

5.2 Data Expansion

부트스트래핑(Bootstrapping)은 전산 언어학에서 레이블이 지정되지 않은 자연언어 데이터의 강도와 레이블이 붙은 데이터의 부족 문제를 해결하기 위하여 사용되었다는 점과 분류되지 않는 데이터와 분류된 데이터 사이의 일반화 오류가 낮을 수 있다는 것을 증명했다는 점에서 관심을 갖게 되었다[6].

부트스트래핑을 통한 데이터 확장에 대한 선행 연구는 기존 데이터보다 작은 사이즈인 N의 크기만큼 데이터를



(그림 2) 데이터 확장 흐름
(Figure 2) Data Expansion Flow

복원 추출하였으며, 이를 반복하여 데이터 크기를 확장하였다. 선행 연구에서는 N은 0~100 사이일 경우로 제한하였으며, N이 15~23 사이일 때 좋은 성능을 보였다[7].

선행 연구를 바탕으로 그림 2과 같은 과정으로 진행하였다. Original Data는 확장 이전의 데이터이다. Class_0는 해파리 출현 지수가 0인 데이터를 뜻하며, Class_1t4는 지수가 1이상인 데이터로 정의하였다. Class_0와 Class_1t4로 클래스를 구분한 이유는 부트스트래핑 간에 비율을 조절하기 위해서이다. r_1, r_2는 각 Class_0, Class_1t4가 Original Data에 차지하는 비율을 나타낸다. 부트스트래핑을 통해 선택할 크기를 T라고 할 때, T를 구하는 공식은 다음과 같다.

$$T = \text{round}(N \times r_1) - 1 \quad \cdots \quad A$$
$$T = \text{round}(N \times r_2) + 1 \quad \cdots \quad B$$

A는 해파리 출현 예측 지수가 0인 Class_0의 부트스트래핑을 통하여 선택할 크기 T를 구하는 식이다. 부트스트래핑으로 만들 작은 사이즈 N에서 Class_0가 차지하는 비율을 곱하고 반올림할 경우 기존 비율보다 높은 수치가 차지하게 된다. 이를 방지하고자 1을 감산해준다. B는 반대로 1을 가산해준다. A, B를 이용하여 얻은 T의 크기만큼 Class_0와 Class_1t4에서 부트스트래핑 과정을 거쳐 데이터를 추출한다. 이 과정을 반복하여 원하는 크기만큼 데이터를 확장한다.

실험에서 N의 크기는 기본으로 15로 하였다. Original Data의 크기는 247이며, 노무라입깃해파리를 기준으로 Class_0의 크기는 209, Class_1t4의 크기는 38이다. r1, r2는 각 84.62, 15.38이다. Class_0에서 부트스트래핑을 통하여 선택할 크기 T는 'round(15*84.62)-1'인 12이며, Class_1t4는 'round(15*15.38)+1'인 3이다. N Size Data의 크기는 15이며, 이를 반복하여 1000, 2000, 10000, 20000의 크기를 갖는 Expanded Data로 확장하였다. 각 해파리 출현 비율은 25%, 36.36%, 15.38%, 15.38%로 증가하였다.

확장 과정을 100회 반복하여 평균, 중앙값, 표본평균을 비교하였으며, 각 확장크기의 모평균에 대한 95% 신뢰구간을 비교하였다. 비교 결과 특징변수의 차이는 표 3에서 보인 임의의 특징변수와 같이 평균, 중앙값, 표본평균 및 신뢰구간에 대해서 큰 차이를 보이지 않았다.

(표 3) Original Data와 Expanded Data 비교
(Table 3) Compare Original and Expanded Data

구분		avr_wind_speed	avr_high_wind_speed
247	평균	4.055667998	6.060565451
	중앙값	4.003970223	5.830208333
	표본평균	0.9088437	1.673207312
1000	평균	4.055675061	6.075834459
	중앙값	3.999657648	5.825161633
	표본평균	0.891120656	1.657544848
	95% CI	3.6190 - 4.4923	5.2636 - 6.8880
2000	평균	4.048527802	6.069459274
	중앙값	3.999192681	5.825993119
	표본평균	0.903661641	1.667978272
	95% CI	3.605 - 4.4913	5.2521 - 6.8867
10000	평균	4.045903814	6.05966464
	중앙값	4.00158214	5.823415081
	표본평균	0.914009507	1.666574479
	95% CI	3.5980 - 4.4937	5.2430 - 6.8762
20000	평균	4.045135848	6.061458926
	중앙값	4.002125184	5.823715803
	표본평균	0.906794537	1.66741902
	95% CI	3.6008 - 4.4894	5.2444 - 6.8784

6. 실험 및 결과

6.1 Collect Data

6.1.1 특징 변수 선정

우리나라에 시행되고 있는 해파리 모니터링 방식이 어업인들의 관측을 통해 이루어진다는 점에서 어업에 영향을 미치는 변수가 해파리 출현 예측에 오류를 범할 가능성을 가질 수 있기에 변수에서 제외하였다. 또한 수집 가능성을 토대로 기온, 기압, 수온, 유속, 풍속, 염분, 조위, 파고 총 8개의 특징 변수를 선정하였다.

6.1.2 데이터 수집

해양 데이터들은 국립해양조사원의 바다누리 해양 정보 서비스를 이용하여 수집하였으며, 해파리 출현 데이터

는 국립수산과학원의 해파리 모니터링 주간보고를 통하여 수집하였다. 이전의 연구에서 노무라입깃해파리, 보름달물해파리, 기타 해파리로 구분하여 지역별 해파리 출현을 예측하였다면, 이번 연구에서는 낮은 출현율을 가진 기타 해파리를 구분하여 출현 예측이 가능함을 확인하고자 노무라입깃해파리, 보름달물해파리, 유령해파리, 기타 해파리로 구분하여 진행하였다.

앞선 SVM을 이용한 부산 연안 해파리 출현 예측 시 광역시/도 단위를 기준으로 97.32%라는 높은 예측률을 보였다. 그러나 높은 예측률의 원인이 너무 넓은 범위를 예측하여 단순한 예측 결과를 도출했다는 점에서 이를 보완하고자 구/군 단위로 범위를 축소하여 해파리 출현을 예측하고자 하였다. 또한 해파리의 일일 이동거리를 고려하여 예측할 구/군과 인접한 구/군을 포함하여 출현을 예측하였다.

실험에서는 우리나라의 가장 많은 해수욕장 이용객 분포를 보이고 있는 해운대 해수욕장을 목표로 하여 해파리 출현 예측을 진행하였다. 예측할 범위는 수영구, 기장군, 해운대구이며, 예측 기간은 2014년 1월부터 2018년 12월까지이다.

6.2 Model Construction

6.2.1 Feature Selection

모델의 예측 정확도를 높일 수 있도록 특징 선택을 진행하였다. 특징 선택 방법은 이전 연구에서 사용된 Random Forest(RF)의 특징 중요도(feature importance)를 측정하는 방법[8]과 간단하고 빠르게 이용할 수 있으며, 역전파 알고리즘을 사용하는 모델에서 좋은 성능을 발휘하는 Sequence Forward Selection(SFS)[9]을 각각 이용하였다.

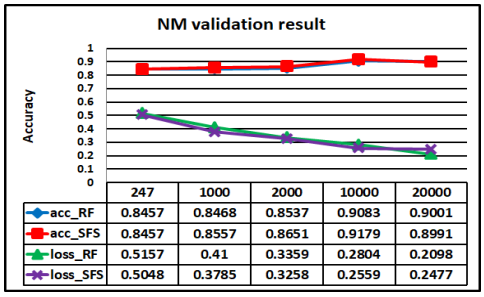
Random Forest를 이용하여 특징 선택 시 26개의 특징 중 24개의 특징이 선택되었으며, Sequence Forward Selection을 이용하여 특징을 선택할 경우 26개 특징 모두 중요성을 갖는 것으로 나타났다. 두 특징 선택 방법이 심층신경망을 이용한 해파리 출현 예측의 정확도에 어떤 영향을 줄 것인지 확인하기 위해 각각 구별하여 실험을 진행하였다.

6.2.2 심층신경망 모델

각 뉴런의 도출된 값은 ReLU 활성함수를 이용하였으며, 손실함수는 Cross entropy를 이용하였다. 은닉 계층의 0.5 확률로 무작위로 생략되도록 하여 데이터 과적합 문제를 보완할 수 있도록 하였다. 최종 0~4까지 5개의 Label으로 분류되도록 하였다.

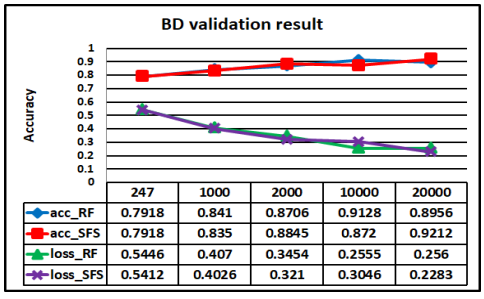
6.2.3 Validation Result

실험 환경은 Windows 10 Education 운영체제를 이용하였으며, Python 3.6버전으로 작성하였다. 모델과 부트스트래핑은 Tensorflow Keras를 이용하여 구현하였다. 검증 방식은 K-Fold 교차 검증을 실시하였으며, 가장 기본적인 10등분으로 나누어 교차 검증을 진행하였다. 교차 검증 결과는 그림 3부터 6과 같다. 전체적으로 노무라입깃해파리, 보름달물해파리, 유령해파리, 기타 해파리의 예측 정확도가 원본 데이터보다 확장된 데이터의 크기가 클수록 높은 수치로 나타났다. 특히, 10000건 이상으로 데이터를 확장할 경우 노무라입깃해파리와 보름달물해파리는 90% 이상의 예측 정확도를 보이고 있으며, 유령해파리와 기타 해파리는 98% 이상의 예측 정확도를 보이고 있다. 특징 선택 방법에 따라 예측률과 손실값 모두 영향을 받았다. 10000건으로 데이터를 확장했을 때, 일부 Random Forest를 이용한 모델이 더 높은 예측률 및 낮은 손실값을 보이고 있으나 전반적으로 Sequence Forward Selection를 이용



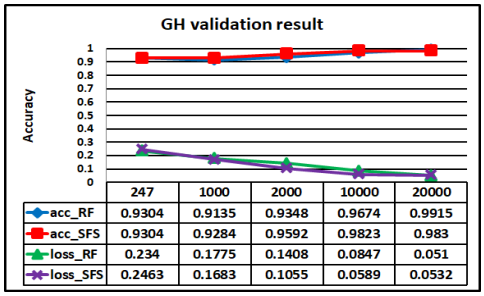
	247	1000	2000	10000	20000
acc_RF	0.8457	0.8468	0.8537	0.9083	0.9001
acc_SFS	0.8457	0.8557	0.8651	0.9179	0.8991
loss_RF	0.5157	0.41	0.3359	0.2804	0.2098
loss_SFS	0.5048	0.3785	0.3258	0.2559	0.2477

(그림 3) 노무라입깃 해파리 검증 결과
(Figure 3) Nemopilema nomurai validation result



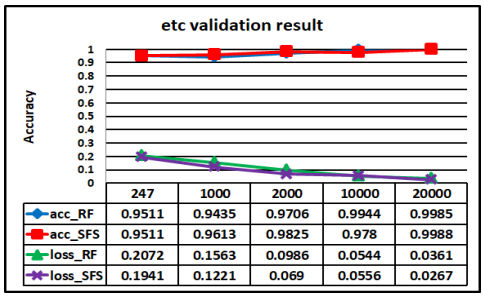
	247	1000	2000	10000	20000
acc_RF	0.7918	0.841	0.8706	0.9128	0.8956
acc_SFS	0.7918	0.835	0.8845	0.872	0.9212
loss_RF	0.5446	0.407	0.3454	0.2555	0.256
loss_SFS	0.5412	0.4026	0.321	0.3046	0.2283

(그림 4) 보름달물 해파리 검증 결과
(Figure 4) Aurelia aurita validation result



	247	1000	2000	10000	20000
acc_RF	0.9304	0.9135	0.9348	0.9674	0.9915
acc_SFS	0.9304	0.9284	0.9592	0.9823	0.983
loss_RF	0.234	0.1775	0.1408	0.0847	0.051
loss_SFS	0.2463	0.1683	0.1055	0.0589	0.0532

(그림 5) 유령 해파리 검증 결과
(Figure 5) Cyanea nozakii validation result



	247	1000	2000	10000	20000
acc_RF	0.9511	0.9435	0.9706	0.9944	0.9985
acc_SFS	0.9511	0.9613	0.9825	0.978	0.9988
loss_RF	0.2072	0.1563	0.0986	0.0544	0.0361
loss_SFS	0.1941	0.1221	0.069	0.0556	0.0267

(그림 6) 기타 해파리 검증 결과
(Figure 6) Other jellyfish validation result

한 모델이 더 좋은 성능을 보였다.

비교를 위하여 적은 데이터에서도 좋은 성능 보이는 Transfer Learning에 원본 데이터를 학습하여 예측률을 비교해보고자 하였다[10]. 실험 방식은 이전 방식과 동일하다. 심층신경망에 Transfer Learning을 접목하여 학습한 결과 표 4와 같은 결과를 얻을 수 있었다. 원본 데이터인 247개의 데이터를 학습하였을 때보다 높거나 또는 비슷한 출현 예측 정확도를 얻을 수 있었으나, 본 논문이 진행한 부트스트래핑을 통하여 데이터를 학습한 결과보다는 약 6%정도 낮은 수치로 나타났다. 특징 선택 방법에 대한 예측 정확도의 차이는 나타나지 않았다.

6.3 Prediction Evaluation

테스트 데이터를 이용하여 각 해파리별 좋은 성능을 보여주고 있는 출현 예측 모델의 예측 결과를 확인하였다. 확인한 결과 노무라입깃해파리 0.8181%, 보름달물해파리 1.0%, 유령해파리 0.63%, 기타 해파리 1.0%의 정확도를 보였다. 유령해파리와 노무라입깃해파리의 예측 정확도가 검증단계에서 확인한 결과보다 낮은 결과를 보였다. 확인한 결과 각 해파리의 예측 정확도와 해파리 출현 지수 0인 해파리가 출현하지 않을 확률과 동일함을 확인하였다.

(표 4) Transfer learning 검증 결과
(Table 4) Transfer learning verification results

구분		RF	SFS
노무라입깃 해파리	Accuracy	0.8578	0.8578
	loss	2.2941	2.2941
보름달물 해파리	Accuracy	0.7918	0.7918
	loss	3.3552	3.3552
유령 해파리	Accuracy	0.9304	0.9304
	loss	1.1202	1.1202
기타	Accuracy	0.9511	0.9511
	loss	0.781	0.781

7. 결 론

해파리 출현 예측 시 수집된 해양 데이터들의 크기가 적다는 한계로 인하여 최대 84.57% 예측률 이상의 성능을 얻을 수 없었다. 이에 본 논문은 부트스트래핑을 이용하여 데이터의 크기 및 해파리 출현 비율을 조절하는 방법으로 해당 문제를 해결하였다. 해파리 출현 비율을 조절 시 데이터 신뢰성에 대해 염려하였으나, 비교한 결과에서 소숫점 한자리 이하의 차이로 큰 차이를 보이지 않았다. 확장된 데이터가 원본 데이터를 학습시킨 결과보다 약 7% 높은 성능의 좋은 결과를 보였으며, Transfer learning과 비교했을 때, 해파리 출현 예측에서는 약 6% 더 좋은 결과를 보였다. 높은 정확도를 보인 예측 모델에 테스트 데이터를 실험한 결과, 노무라입깃해파리와 유령해파리의 경우 검증한 결과보다 낮은 결과를 보였다. 해당 결과는 해파리가 출현하지 않는 해파리 출현 지수 0의 비율과 동일함을 확인하였다. 결론적으로 해파리의 출현 유무를 예측하는 것은 높은 정확도로 가능하다는 것을 확인하였으나, 해파리의 출현을 5단계로 지수화하여 예측하는 것은 아직 낮은 성능을 보임을 확인할 수 있었다.

향후 과제로 해파리의 출현에 대한 지수화를 통하여 해수욕장 해파리 쏘임 사고 예방 및 기타 피해를 최소화할 수 있도록 연구되어야 한다.

참고문헌(Reference)

[1] Kim, Dae-Young, Lee, Jung-Sam, Kim, Do-Hoon, “A Study on Direction of Industrial Utilization for Jellyfish in Korea,”The Korean Society for Fisheries and Marine Sciences Education, Volume 26, Issue 3, pp.587-596, 2014.
https://doi.org/10.13000/JFMSE.2014.26.3.587

[2] KIM, Bong-Tae, EOM, Ki-Hyuk, HAN, In-Seong, PARK, Hye-Jin, “An Analysis of the Impact of Climatic Elements on the Jellyfish Blooms,” The Korean Society for Fisheries and Marine Sciences Education, Volume 27, Issue 6, pp.1755-1763, 2015, https://doi.org/10.13000/JFMSE.2015.27.6.1755

[3] T. Nishikawa, K. Miyahara, T. Ohtani, T. Senjyu, “Occurrence and potential prediction of the giant jellyfish Nemopilema nomurai off Hyogo Prefecture, southwestern Sea of Japan, during 2006–2015,” Regional Studies in Marine Science, Vol.16, pp.181-187, 2017.
https://doi.org/10.1016/j.rsma.2017.09.002

[4] R. Mu, X. Zeng, “A Review of Deep Learning Research,” KSII Transactions on Internet and Information Systems, Vol.13, No.4, pp.1738-1764, 2019.
https://doi.org/10.3837/tiis.2019.04.001

[5] M. C. Acock, Y. A. Pachepsky, “Estimating missing weather data for agricultural simulations using group method of data handling,” Journal of Applied meteorology, Vol.39, No.7, pp.1176-1184, 2000.
https://doi.org/10.1175/1520-0450(2000)039<1176:emwdfa>2.0.co;2

[6] S. Abney, “Bootstrapping,” Proceedings of the 40th Annual Meeting on Association for Computational Linguistics Association for Computational Linguistics, pp.360-367, 2002.

[7] R. B. Akins, H. Tolson, B. R. Cole, “Stability of response characteristics of a Delphi panel: application of bootstrap data expansion,” BMC medical research methodology, Vol.5, No.1, pp.37-49, 2005.
https://doi.org/10.1186/1471-2288-5-37

[8] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, F. A. Hamprecht, “A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data,” BMC bioinformatics, Vol.10, No.1, pp.213-229, 2009.
https://doi.org/10.1186/1471-2105-10-213

[9] A. Marcano-Cedeno, J. Quintanilla-Dominquez, M. G. Cortina-Januchs, D. Andina, “Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network,” In IECON 2010-36th annual conference on IEEE industrial electronics society, pp.2845-2850. 2010.
https://doi.org/10.1109/iecon.2010.5675075

[10] S. J. Pan, Q. Yang, “A survey on transfer learning,” IEEE Transactions on knowledge and data engineering, Vol.22, No.10, pp.1345-1359, 2010.
https://doi.org/10.1109/tkde.2009.191

◐ 저 자 소 개 ◑

황 철 훈(CHEOLHUN HWANG)
2019년 가천대학교 컴퓨터공학과(공학사)
2019년~현재 가천대학교 일반대학원 소프트웨어학과(공학석사)
관심분야 : 정보보호, 머신러닝, 인공지능, 모바일
E-mail : qewqsa@naver.com

한 명 묵(Myung-Mook Han)
1980년 연세대학교 공과대학(공학사)
1987년 뉴욕공과대학교 대학원 컴퓨터공학과(이학석사)
1997년 오사카시립대학교 대학원 정보공학부(공학박사)
1998년~현재 가천대학교 소프트웨어학과 교수
관심분야 : 정보보호, 알고리즘, 데이터 마이닝
E-mail: mmhan@gachon.ac.kr