

# 순환신경망을 이용한 뜰개의 관측 데이터 보정

김경도, 김용혁\*  
광운대학교 컴퓨터과학과

## Correction of Drifter Data Using Recurrent Neural Networks

Gyoung-Do Kim, Yong-Hyuk Kim\*  
Dept. Computer Science, Kwangwoon University

요 약 해양 뜰개는 해수면을 떠다니며 해양 기상 등을 관측하는 장비로, 뜰개를 통해 관측한 데이터는 해양 기상 예측, 유류유출 예측 등의 상황에서 활용된다. 관측 데이터는 관측 시에 오측(error data) 또는 결측(missing data)이 발생할 수 있으며, 오측 또는 결측된 데이터가 포함 될 경우, 데이터를 사용하는 모델들의 정확도가 떨어질 수 있다. 본 논문에서는 데이터 보정을 위한 방법으로 순환신경망을 이용한 데이터 보정 모델을 제안한다. 2015년 7개, 2016년 8개의 뜰개를 통해 수집한 해양 데이터를 이용한 보정 실험 결과와 보정 결과를 검증하기 위한 뜰개 이동 예측 실험을 설명하며, 실험 결과, 데이터 보정을 통해 13.9%의 데이터가 보정되었으며, 이동 예측 모델의 성능이 1.4% 향상되는 것을 보였다.

주제어 : 데이터 보정, 순환신경망, 기계학습, 예측, 뜰개

**Abstract** The ocean drifter is a device for observing the ocean weather by floating off the sea surface. The data observed through the drifter is utilized in the ocean weather prediction and oil spill. Observed data may contain incorrect or missing data at the time of observation, and accuracy may be lowered when we use the data. In this paper, we propose a data correction model using recurrent neural networks. We corrected data collected from 7 drifters in 2015 and 8 drifters in 2016, and conducted experiments of drifter moving prediction to reflect the correction results. Experimental results showed that observed data are corrected by 13.9% and improved the performance of the prediction model by 1.4%.

**Key Words** : Data Correction, RNN, Machine Learning, Prediction, Drifter

### 1. 서론

해양 뜰개(Drifter)는 해수면을 떠다니며 해양 기상 등을 관측하는 장비로서 사용되며, 모습은 Fig. 1 과 같다. 해양 뜰개를 이용하여 해류 및 바람의 흐름과 세기 등을 측정 가능하며, 이러한 관측 데이터는 기상 예측 혹은 유류 유출 등의 상황에서 활용이 가능하다[1]. 하지만, 이러한 데이터 수집 과정에서 주변 환경의 영향 및 기계적 결함 등으로 인해 측정 범위를 벗어나는 오측(error data)

과 데이터를 제대로 측정하지 못하는 결측(missing data)이 발생 할 수 있으며, 잘못된 데이터를 이용해 실험 또는 모델 학습을 진행할 경우, 잘못된 실험 결과나 모델의 성능을 하락시키는 원인이 된다[2-4]. 부정확한 데이터를 제거하고 정확한 데이터를 이용해 실험을 진행하기 위해서는 데이터 보정 과정이 필요하다. 데이터 보정을 위해 다양한 기계학습을 이용한 연구들이 진행되었으며[5-7], 기계학습 기반의 보정 모델을 통해 기상 관측 데이터들의 보정을 진행할 경우 예측 정확도가 증가하였다[8,9].

\*This research was supported by a grant [KCG-01-2017-05] through the Disaster and Safety Management Institute funded by Korea Coast Guard of Korean government.

\*Corresponding Author : Yong-Hyuk Kim (yhdhfly@kw.ac.kr)

Received January 26, 2018

Revised February 26, 2018

Accepted March 20, 2018

Published March 28, 2018



Fig. 1. An ocean drifter to observe ocean weather phenomena

본 논문에서는 15개의 뜰개를 이용해 관측한 데이터를 순환신경망 기반의 보정 모델을 통해 보정하는 방법을 제안한다. 본 논문의 2 절에서는 실험에서 사용한 실험 데이터의 특징, 속성 등을 설명한다. 3 절에서는 본 논문에서 제안하는 보정 모델에 사용된 기법인 순환신경망에 대해 소개한다. 4 절에서는 제안하는 보정 모델을 소개하며, 보정 모델을 이용해 진행된 보정 실험 및 실험 결과를 소개한다. 5 절에서는 보정 모델을 통해 보정한 데이터를 이용해 뜰개 이동 예측 실험을 진행한 결과를 소개하며, 마지막 6 절에서는 보정 실험에 대한 결론 및 이를 토대로 한 향후 연구방향을 제시한다.

Table 1. Features of test data

Location	Measured period	Number of drifters	Data attribute	Number of data
120 km distant from the west coast of Tae-an-gun	10 days in November 2015	7	Time, latitude, longitude, wind, ocean current, etc. Total 12 values	683
10 km distant from the west coast of Jeju-Island	5 days in April 2016	8		710

## 2. 실험 데이터

본 연구에서 사용한 실험 데이터는 2015년 서해상에서 7개의 뜰개를 이용해 약 10 일간 한 시간 단위로 관측한 683개의 데이터와 2016년 제주도 서해상에서 8개의 뜰개를 이용해 약 5 일간 관측한 데이터 710개를 사용하였으며, 각각의 데이터는 뜰개가 데이터를 측정할 시간과 관측한 바람( $wind_u$ ,  $wind_v$ )과 해류( $current_u$ ,  $current_v$ ) 측정 값, 그리고 측정 당시 뜰개의 위치를 위도와 경도 값으로 측정된 값으로 구성되어 있다.

Table 1은 실험에 사용한 데이터의 특징을 나타냈으며, Table 2는 데이터의 속성의 종류와 설명을 나타낸 표이다.

Table 2. Attributes of experiment data

Attribute	Description
longitude	Longitude value of the drifter position
latitude	Latitude value of the drifter position
$wind_u$	Wind value observed by the drifter
$wind_v$	Wind value observed by the drifter
$current_u$	Current value observed by the drifter
$current_v$	Current value observed by the drifter

## 3. 순환신경망

순환신경망은 딥러닝의 한 종류로서 딥러닝은 여러 비선형 변환의 조합을 통해 높은 수준의 추상화 모델을 구축하는 기계학습 기법이다. 딥러닝은 음성인식, 자연어 및 음성, 신호처리 등의 다양한 분야에서 뛰어난 성능을 보여주고 있다. 그 중 순환신경망은 기상 관측 데이터와 같은 순차적인 정보를 처리하는데 초점을 맞춘 신경망으로 시계열 형태의 데이터 구조에서 뛰어난 성능을 보여준다[10,11]. 순환신경망은 아래 Fig. 2와 같이 신경망이

반복되는 구조로 되어 있으며, 다른 신경망 구조에서는 모든 입력과 출력이 각각 독립적이라고 가정하지만 순환 신경망에서는 입력으로 이전 단계의 출력 결과를 함께 사용함으로써, 과거의 데이터를 현재 혹은 미래의 문제 해결에 활용할 수 있는 구조를 가지고 있다.

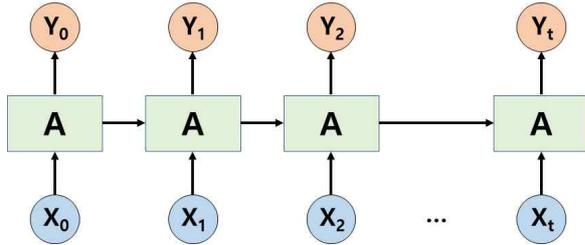


Fig. 2. Structure of recurrent neural network

#### 4. 데이터 보정 실험

본 절에서는 논문에서 제안하는 데이터 보정 모델에 대해 소개하고, 해당 모델을 이용해 보정 실험 및 실험 결과를 소개한다.

##### 4.1 데이터 보정 모델

본 논문에서 제안하는 데이터 보정 모델은 순환신경망 기반의 모델로, 정상적인 데이터를 이용해 순환신경망 모델을 학습시킨 후 첫 번째부터 정해진 개수까지 데이터의 표준편차를 구하고, 해당 표준편차를 변화량의 한계점으로 설정한다. 그 다음, 입력받은 관측 데이터를 모델에 입력시켜 입력받은 데이터의 변화량이 한계점을 벗어날 경우, 데이터 보정을 진행하는 방식으로 동작한다. 데이터 보정 모델의 파라미터 설정으로는 셀의 크기는 10개, 학습률은 0.01, 학습 반복 횟수는 700회, 옵티마이저는 AdamOptimizer를 사용하였다. 아래 알고리즘 1

은 제안하는 데이터 보정 모델의 의사코드이며, Fig. 3은 데이터 보정 모델의 흐름도이다.

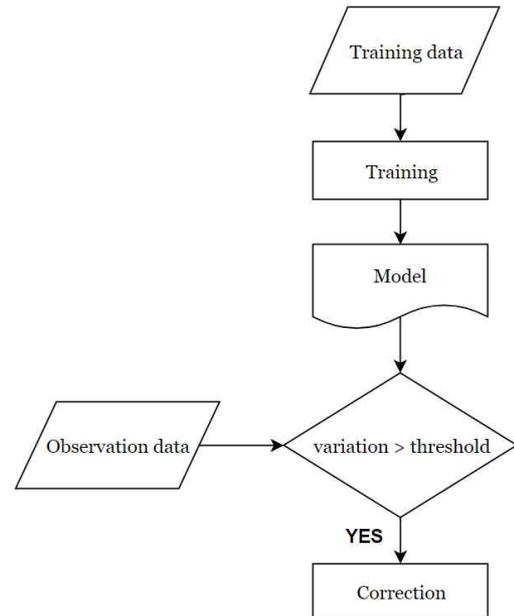


Fig. 3. Flowchart of data correction model

##### 4.2 데이터 보정 실험 및 결과

데이터 보정 실험은 관측한 뜰개의 데이터 중 불규칙적인 변화가 많은 바람 측정 값인  $wind_u$ ,  $wind_v$  데이터 속성을 대상으로 진행하였으며, 전체 7개 케이스의 뜰개 데이터 중, 데이터가 많은 1,5,6,7 번 케이스의 데이터를 이용해 보정 실험을 진행하였다. 실험 결과, 평균 13.9%의 데이터가 보정이 되었으며, 보정된 데이터의 변화는 Fig. 4와 Fig. 5와 같다. 그래프들은 데이터들의 수치를 나타냈으며, 주황색 실선은 기존 데이터의 수치 값을, 파란색 실선은 보정된 데이터의 수치 값을 표현했다. Table 3은 케이스 별 데이터 보정 결과를 나타낸 표이다.

---

#### Algorithm 1 Correction

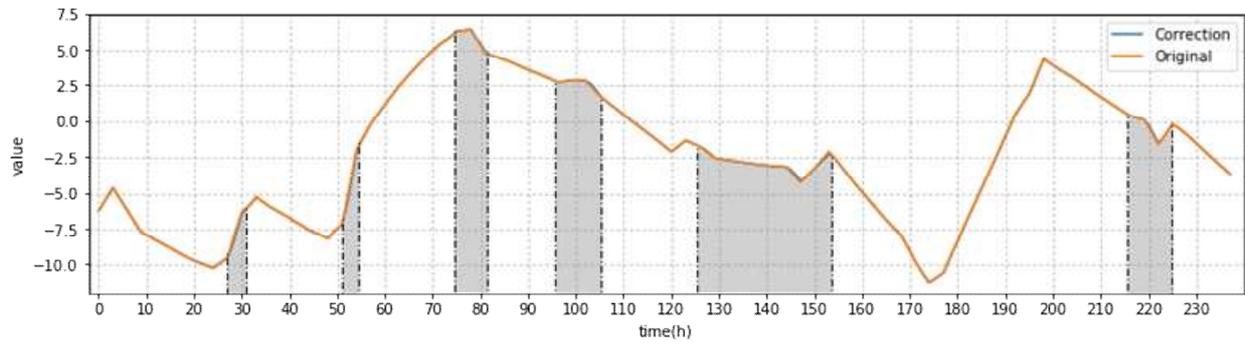
---

```

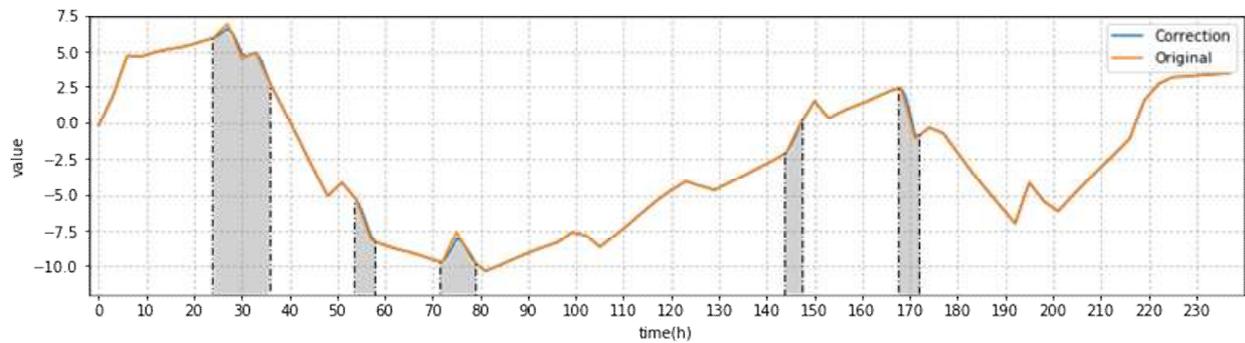
1: procedure CORRECTION(ob_datai from ob_data0 to i-1)
2:   model ← Load an RNN model trained through ob_data;
3:   thresholdsi ← Standard deviation(ob_datai to i-k);           ▷ we set k to 10.
4:   variationi ← ob_datai - ob_datai-1;
5:   if variationi > thresholdsi then
6:     pred_val ← Predict using model(ob_datai);
7:     ob_datai ← (ob_datai + pred_val)/2;
8:   end if
9: end procedure

```

---

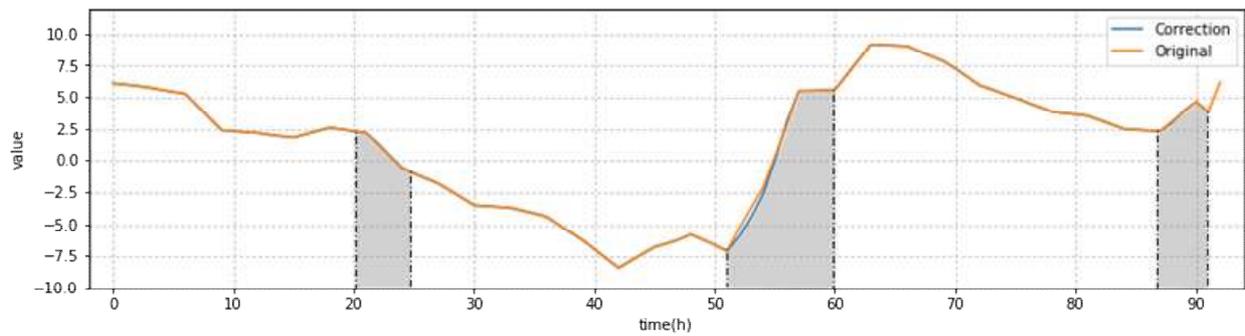


(a) wind<sub>u</sub> in Case 1-1

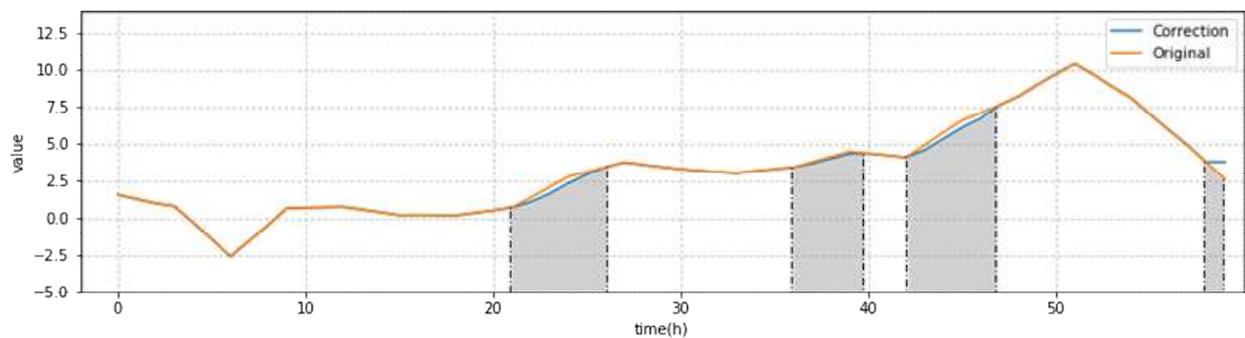


(b) wind<sub>v</sub> in Case 1-1

Fig. 4. An example result of wind correction in Case 1



(a) wind<sub>u</sub> in Case 2-8



(b) wind<sub>v</sub> in Case 2-5

Fig. 5. An example result of wind correction in Case 2

Table 3. Statistics of corrected data

Case	Number of data	Number of correction data		Correction rate (%)
		wind <sub>u</sub>	wind <sub>v</sub>	
1-1	238	24	23	9.9
1-5	108	18	8	12.0
1-6	111	12	16	12.6
1-7	112	14	14	12.5
2-1	82	13	13	15.9
2-2	106	16	16	15.1
2-3	106	18	18	17.0
2-4	87	12	12	13.8
2-5	60	12	12	20.0
2-6	93	12	12	12.9
2-7	94	12	12	12.8
2-8	93	11	11	11.8

### 5. 뜰개 이동 예측 실험

본 절에서는 보정한 데이터를 이용해 진행한 뜰개 이동 예측 실험 및 실험 결과를 소개한다. 뜰개 이동 예측 실험은 유류유출 및 기상 예측에서 뛰어난 예측 성능을 보여준 기계학습 기법들을 선정하여[12], 뜰개가 관측한 시간별 해류와 바람의 값을 데이터로 사용하여 시간별 위도, 경도 변화량을 진화연산[13] 또는 기계학습 모델을 이용해 예측하는 실험이다[14,15]. 해당 실험은 유류 유출 확산 예측 연구의 선행 연구로서 진행되었으며, 실시간으로 측정 가능한 해류와 바람의 관측 값을 이용해 해상에서의 이동 방향을 예측하기 위한 목적으로 진행되었다. 본 논문에서는 기존 관측 데이터와 보정을 진행한 관측 데이터를 이용해 뜰개 이동 예측 연구를 진행하였으며, 데이터 보정이 예측 모델의 정확도를 향상시키는데 기여했는지를 검증하기 위한 방법으로 사용하였다.

#### 5.1 뜰개 이동 예측 모델

본 논문에서 실험에 사용한 뜰개 이동 예측 모델은 순환신경망을 이용하였으며, 모델의 입력으로는 뜰개가 관측한 바람과 해류의 데이터 속성 4 가지(wind<sub>u</sub>, wind<sub>v</sub>, current<sub>u</sub>, current<sub>v</sub>)를 입력 값으로 사용하였으며, 출력 값은 모델이 예측한 위도와 경도의 변화량이 출력되도록 하였다.

#### 5.2 뜰개 이동 예측 실험 및 결과

뜰개 이동 예측 실험은 전체 케이스 데이터 중 예측을

진행하는 케이스의 데이터를 테스트 데이터로 사용하고 그 외의 케이스 데이터를 학습 데이터로 사용하여 모델을 학습시켜 실험을 진행하였다. 실험 결과 검증을 위해 실제 뜰개가 이동하며 관측한 위도, 경도 값과 모델이 예측한 위도, 경도 값의 차이를 평균 절대 오차(Mean absolute error, MAE)로 나타내어 비교하였다. Table 4 보정을 진행한 wind<sub>u</sub>, wind<sub>v</sub> 속성의 케이스별 평균 보정 수치를 정리한 표이며, Table 5는 기존 데이터를 이용한 뜰개 이동 예측 결과와 보정된 데이터를 이용한 뜰개 이동 예측 결과를 케이스 별 평균 절대 오차 값으로 정리한 결과이다. Fig. 6과 Fig. 7은 실제 뜰개가 이동한 경로와 뜰개 이동 예측 모델을 이용하여 예측한 결과를 비교한 그림이며, 경로가 보정된 부분을 빨간선으로 표시하였다.

Table 4. Average of difference between original value and corrected one

Case	Correction value	
	wind <sub>u</sub>	wind <sub>v</sub>
1-1	0.0021	0.0058
1-5	0.0149	0.0006
1-6	0.0031	0.0043
1-7	0.0059	0.0070
2-1	0.0102	0.0088
2-2	0.0067	0.0065
2-3	0.0017	0.0118
2-4	0.0140	0.0057
2-5	0.0506	0.0576
2-6	0.0013	0.0039
2-7	0.0066	0.0139
2-8	0.0197	0.0131

Table 5. Prediction results of drifter movement

Case	Original result	Corrected result	MAE	Improvement rate
1-1	0.0704	0.0697	0.0007	1.0%
1-5	0.0732	0.0728	0.0004	0.5%
1-6	0.0271	0.0266	0.0005	1.8%
1-7	0.0647	0.0650	-0.0003	-0.5%
2-1	0.0471	0.0468	0.0003	0.6%
2-2	0.0599	0.0583	0.0016	2.7%
2-3	0.0286	0.0286	0.0000	0.0%
2-4	0.0519	0.0514	0.0005	1.0%
2-5	0.0493	0.0491	0.0002	0.4%
2-6	0.0282	0.0276	0.0006	2.1%
2-7	0.0335	0.0323	0.0012	3.6%
2-8	0.0532	0.0516	0.0016	3.0%
평균				1.4%

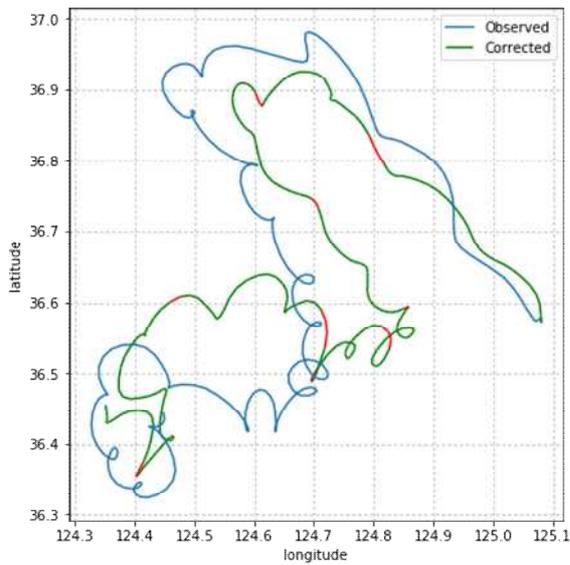


Fig. 6. An example result of drifter movement prediction in Case 1-1

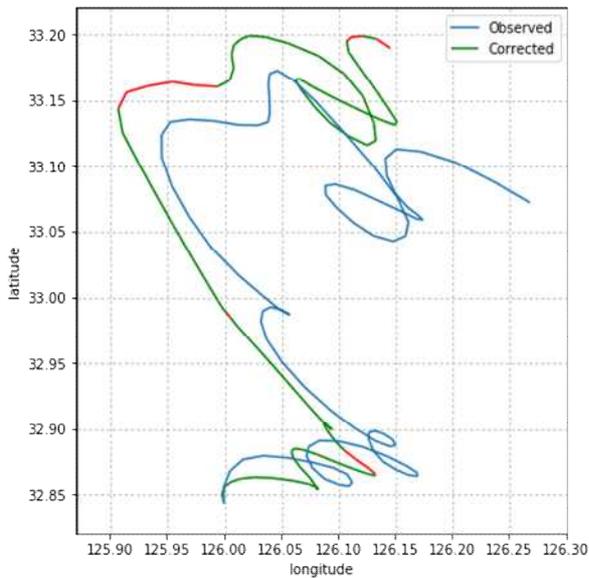


Fig. 7. An example result of drifter movement prediction in Case 2-7

## 6. 결론

본 논문에서는 뜰개를 이용해 관측한 데이터에서 발생 가능한 오측 또는 결측을 해결하기 위해 순환신경망을 이용한 데이터 보정 모델을 제안하였다. 순환신경망 기반 데이터 보정 모델을 통해 관측 데이터의 평균 13.9%를 보정하였으며, 보정된 데이터를 이용한 예측 모델의 성능 또한 1.4% 향상되는 결과를 보였다. 연구 결과

를 통해 노이즈 값을 가지고 있는 시계열 데이터의 보정 가능성을 제시하였으며, 데이터 보정을 통해 데이터를 이용하는 예측 모델의 성능 또한 향상시킬 수 있다는 결과를 제시하였다.

본 논문에서는 향후 연구 방향으로 더 많은 데이터와 다양한 파라미터를 입력으로 사용하는 보정 연구를 제시하며, 오토인코더(Autoencoder) 기법을 이용한 데이터 보정 방안을 제시한다.

## REFERENCES

- [1] J. S. Park, K. R. Kang, S. Lee & S-R. Lee. (2013). Observation of the Sea Surface Skin Current Using a GPS-Drifter. *Ocean and Polar Research*, 35(4), 193-203. DOI : 10.4217/OPR.2013.35.3.193
- [2] R. J. Little & D. B. Rubin. (2014). *Statistical analysis with missing data*. John Wiley & Sons. DOI : 10.1002/9781119013563
- [3] J. W. Graham. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology*. *Annual Review of Psychology*, 60, 549-576. DOI : 10.1146/annurev.psych.58.110405.085530
- [4] N. F. Schneidewind & H. M. Hoffmann, (1979). An experiment in software error data collection and analysis. *IEEE Transactions on Software Engineering*, 3, 276-286. DOI : 10.1109/tse.1979.234188
- [5] J. H. Ha & Y. H. Kim, (2016). A Survey on Data Correction of Observation and Prediction Using Machine Learning: Preliminary Study for Optimizing Oil Spill Model. *International Journal of Applied Engineering Research*, 11(16), 8892-8895. DOI : 10.6088/ijaser.
- [6] D. E. Rumelhart, G. E. Hinton & R. J. Williams. (1985). Learning Internal Representations by Error Propagation. *DTIC Document, Tech. Rep.* DOI : 10.21236/ada164453
- [7] J. C. Platt. (1999). Fast Training of Support Vector Machines Using Sequential Minimal Optimization. *Advances in Kernel Methods*. 185-208.
- [8] M. K. Lee, S. H. Moon, Y. H. Kim & B. R. Moon. (2014). Correcting Abnormalities in Meteorological Data by Machine Learning. *In Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference*, 888-893. DOI : 10.1109/smc.2014.6974024

- [9] Y. H. Kim, J. H. Ha, Y. Yoon, N. Y. Kim, H. H. Im, S. Sim & R. K. Choi. (2016). Improved Correction of Atmospheric Pressure Data Obtained by Smartphones through Machine Learning. *Computational intelligence and neuroscience*, 2016(4).  
DOI : 10.1155/2016/9467878
- [10] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis & P. S. Dokopoulos. (2006). Long-term wind speed and power forecasting using local recurrent neural network models. *IEEE Transactions on Energy Conversion*, 21(1), 273-284.  
DOI : 10.1109/tec.2005.847954
- [11] T. Guo, Z. Xu, X. Yao, H. Chen, K. Aberer & K. Funaya. (2016). Robust Online Time Series Prediction with Recurrent Neural Networks. *In Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference*, 816-825.  
DOI : 10.1109/dsaa.2016.92
- [12] G. D. Kim & Y. H. Kim, (2017). A Survey on Oil Spill and Weather Forecast Using Machine Learning Based on Neural Networks and Statistical Methods. *Journal of the Korea Convergence Society*, 8(10), 1-8.  
DOI : 10.15207/JKCS.2017.8.10.001
- [13] Y. W. Nam & Y. H. Kim, (2018). Prediction of Drifter Trajectory Using Evolutionary Computation. *Discrete Dynamics in Nature and Society 2018, 2018, In press*.
- [14] C. J. Lee, G. D. Kim & Y. H. Kim, (2017). Performance Comparison of Machine Learning Based on Neural Networks and Statistical Methods for Prediction of Drifter Movement. *Journal of the Korea Convergence Society*, 8(10), 45-52.  
DOI : 10.15207/JKCS.2017.8.10.045
- [15] C. J. Lee & Y. H. Kim, (2017). Prediction of Drifter Trajectory Using an Ensemble Technique. *Convergence Research Letter*, 3(4), 671-674.

김 경 도(Kim, Gyoung Do)

[정회원]



- 2016년 8월 : 광운대학교 컴퓨터 소프트웨어전공 학사
- 2016년 9월 ~ 현재 : 광운대학교 컴퓨터과학과 석사과정
- 관심분야 : 데이터마이닝, 기계학습, 최적화

▪ E-Mail : rilyaus@kw.ac.kr

김 용 혁(Kim, Yong Hyuk)

[정회원]



- 1999년 2월 : 서울대학교 전산과학전공 학사
- 2001년 2월 : 서울대학교 전기컴퓨터공학부 석사
- 2005년 2월 : 서울대학교 전기컴퓨터공학부 박사

- 2005년 3월 ~ 2007년 2월 : 서울대학교반도체공공동원구소 연구원
- 2007년 3월 ~ 2017년 2월 : 광운대학교 컴퓨터소프트웨어학과 조교수/부교수
- 2017년 3월 ~ 현재 : 광운대학교 소프트웨어학부 교수
- 관심분야 : 최적화, 진화연산, 지식공학
- E-Mail : yhdfly@kw.ac.kr