

# 돌발홍수 예보를 위한 빅데이터 분석방법

박다인\*, 윤상후\*\*

대구대학교 통계학과 석사과정\*, 대구대학교 전산통계학과 조교수\*\*

## The big data method for flash flood warning

Dain Park\*, Sanghoo Yoon\*\*

Dept. of Statistics, Daegu University\*

Dept. of Computer Science and Statistics, Daegu University\*\*

**요 약** 돌발홍수는 강우유출수가 하천으로 모여드는 유역이 좁은 지역에 집중호우로 인해 유입되는 물의 양이 급증하여 나타난다. 돌발홍수는 유속이 빠르고 홍수를 대비할 수 있는 시간이 부족하므로 인명과 재산상의 피해를 발생시킨다. 본 연구에서는 돌발홍수를 예보를 위한 빅데이터 분석방법을 수행하였다. 연구 자료는 2009년에서 2012년까지 국민안전처 국가재난정보센터에 보고된 38건의 홍수 피해 자료와 지표수문모형(TOPLATS)에 의해 생성된 수문기상정보인 강우량, 토양수분 상태, 지표유출량이다. 돌발홍수 발생 선행 6시간의 강우량, 토양수분 상태, 지표유출량 데이터를 요인분석을 통해 토양수분 상태, 장기요인에 의한 강우량과 지표유출량, 단기요인에 의한 강우량과 지표유출량으로 축소하였다. 빅데이터 분석 방법으로는 유형분석인 의사결정나무, 랜덤포레스트, 나이브베이즈, 서포트벡터머신, 로지스틱 회귀모형을 사용하였다. 돌발홍수 사고발생 자료가 38건으로 한정되어 있기 때문에 예측성능 정확도 판단이 중요하다. 예측성능 정확도 평가 방법으로 kappa계수, TP Rate, FP Rate, F-Measure를 이용하였다. 이 외에 돌발홍수 발생 선행 시점별 재현성 평가와 과거 4년간 돌발홍수 경보 횟수를 통해 최적 유형분석 방법을 제시하였다. 연구결과 로지스틱회귀모형과 랜덤포레스트가 돌발홍수 예보를 위한 예측 성능이 가장 좋았다. 사고발생 자료가 2009년부터 2012년까지 38건으로 한정되어 있어 분석을 위한 훈련자료와 검증자료 구축에 한계가 있었다. 장기간의 자료가 수집된다면 더욱 정확한 빅데이터 분석을 수행할 수 있다.

**주제어** : 돌발홍수, 로지스틱회귀모형, 머신러닝, 예보시스템, 지표수문모형.

**Abstract** Flash floods is defined as the flooding of intense rainfall over a relatively small area that flows through river and valley rapidly in short time with no advance warning. So that it can cause damage property and casualty. This study is to establish the flash-flood warning system using 38 accident data, reported from the National Disaster Information Center and Land Surface Model(TOPLATS) between 2009 and 2012. Three variables were used in the Land Surface Model: precipitation, soil moisture, and surface runoff. The three variables of 6 hours preceding flash flood were reduced to 3 factors through factor analysis. Decision tree, random forest, Naive Bayes, Support Vector Machine, and logistic regression model are considered as big data methods. The prediction performance was evaluated by comparison of Accuracy, Kappa, TP Rate, FP Rate and F-Measure. The best method was suggested based on reproducibility evaluation at the each points of flash flood occurrence and predicted count versus actual count using 4 years data.

**Key Words** : Flash flood, Land Surface model, Logistic regression model, Nature hazard warning ,Machine learning

Received 28 September 2017, Revised 30 October 2017  
Accepted 20 November 2017, Published 28 November 2017  
Corresponding Author: Sanghoo Yoon  
(Dept. of Computer Science and Statistics, Daegu University)  
Email: statstar@daegu.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. 서론

최근 지구온난화, 엘니뇨, 라니냐 등 이상기후에 따른 집중호우의 증가로 돌발홍수 피해가 늘어나고 있다. 1970년 이후 월 강우량을 비교분석한 결과 한반도 동부와 남동부는 8월 강우량이 증가했고, 북부와 중서부는 6월 강우량이 증가했다[1]. 1973년부터 2007년까지 홍수기와 비홍수기로 나누어 집중호우빈도를 분석한 결과 홍수기 남한의 연평균강우량이 65.2% 증가하였다[2]. 집중호우빈도의 증가는 한반도지역의 돌발홍수와 연관성이 있는 것으로 판단된다[3].

돌발홍수는 강우유출수가 하천으로 모여드는 유역의 지형이 100km<sup>2</sup> 이내로 좁고 경사가 급한 곳이나 동일한 국지지역 내의 짧은 집중호우로 인해 발생한다. 돌발홍수는 단시간에 하천수위를 상승시키므로 인명과 재산상의 피해를 초래할 수 있다. 한국형 돌발홍수 예보시스템(Flash Flood Guidance, FFG)개발을 위해 수문기상학적 이론 배경이 제시되었고, 프로메티와 엔트로피 기법을 이용한 돌발홍수 위험도 평가 기법이 제안되었으며, TOPPLATS 모형 기반 정량적 돌발홍수 위험성 평가를 통한 돌발홍수를 예·경보를 위한 기술개발 연구가 수행되었다[4,5,6].

돌발홍수는 피해지역의 유형에 따라 도시돌발홍수, 산지돌발 홍수로 표현된다[7]. 돌발홍수 예보를 위한 수문기상정보는 강우량, 지표유출량, 토양수분 상태가 기본적인 모형이다[8]. 지표유출량이란 토양수분이 포화된 상태에서 특정유역의 소하천 제방이 월류하는데 필요한 강우량으로 정의되며, 토양수분 상태는 강우량 및 증발산량 정도에 따라 변하는 값이다. 세계기상기구(WMO)는 돌발홍수를 6시간 이내에 건조한 지표상태에서 집중호우로 인해 하천의 수위가 위험수위까지 상승하고 제방이 월류하는 등, 홍수징후가 시작되는 사례로 정의하고 있다.

통계적 돌발홍수지수는 돌발홍수 피해사례와 지표수문모형으로부터 생산된 돌발홍수 발생 선행 6시간 강우량, 지표유출량, 토양수분상태의 일차선형함수식으로 표현된다[6]. 통계적 돌발홍수지수는 돌발홍수 발생가능성을 수치적으로 표현하므로 실시간 모니터링이 가능하다는 장점이 있으나 돌발홍수의 발생여부 판단의 정확성에서는 한계가 있다. 따라서 본 연구에서는 돌발홍수의 발생여부 판단에 중점을 두어 빅데이터 분석을 수행하였다.

본 연구의 관심지역은 산지지역에서 발생한 돌발홍수이다. 도심지역에서 발생하는 돌발홍수는 침수피해를 발생시키나 산지지역의 돌발홍수는 등산객이나 야영객의 고립과 같은 인명 및 재산상의 피해를 유발하기 때문이다. 분석을 위한 자료는 국가재난정보센터에 보고된 2009년부터 2012년까지 수도권 지역(서울, 경기, 인천)에서 돌발홍수로 인해 발생한 구조요청 사례 38건이다. 또한 돌발홍수 모의를 위한 외부요인으로 지표수문모형에 의해 생성된 지표수문기상정보(강우량, 지표유출량, 토양수분 상태)가 이용되었다.

## 2. 빅데이터 분석 방법

이번 절에서는 돌발홍수 예보를 위한 빅데이터 분석 기법에 대해 설명하고자 한다. 빅데이터 분석 방법으로 의사결정나무, 랜덤포레스트, 나이트베이즈, 서포트벡터머신, 로지스틱 회귀모형을 사용하였다.

### 2.1 의사결정나무(Decision tree)

의사결정나무는 의사결정규칙을 도표화하여 관심대상이 되는 집단을 몇 개의 소집단으로 분류 하거나 예측하는 분석방법이다. 분석과정이 나무구조에 의해서 표현되기 때문에 분석과정을 쉽게 이해하고 설명할 수 있다. 의사결정나무 구조는 뿌리노드로 시작하고 노드분리는 분리 대상인 노드를 자녀 노드로 나눈다. 변수와 분리 값의 선택은 노드의 불순도로 결정한다[9, 10]. 불순도란 집합에 이질적인 것이 얼마나 섞였는지를 측정하는 지표이다. 각 마디에서 분리변수와 분리 기준은 자식마디의 불순도가 가장 큰 분리변수와 분리기준이다. 정지규칙에 의해 현재 마디에서 더 이상 분리가 일어나지 않으면 가지치기 단계로 넘어가 부적절하거나 불필요한 가지를 제거한다. 본 연구에서는 R의 'rpart'패키지를 이용하여 분류와 회귀에 적용되는 CART(Classification And Regression Tree) 분석을 수행하였다[11].

### 2.2 랜덤포레스트(Random Forest)

랜덤포레스트는 여러 개의 의사결정나무를 붓스트랩으로 생성하여 각각의 결과를 앙상블 방법(ensemble methods)으로 종합한다. 붓스트랩으로 생성된 각각의 의

사결정나무는 독립이므로 의사결정나무의 수가 많아도 랜덤포레스트는 과적합 하지 않는다는 장점이 있다[12]. 본 연구에서는 R의 randomForest' 패키지를 이용하였다[13].

### 2.3 나이브베이즈 (Naive Bayes)

나이브베이즈는 베이즈정리를 이용하여 분류하고자 하는 대상의 각 분류별 확률을 측정하여 그 확률이 큰 쪽으로 분류하는 기법이다. 분류될 객체의 종속변수를 나타내는 벡터는  $\mathbf{x}$ 로 표현되며, 나이브베이즈 분류기는 이 벡터를 이용하여  $k$ 개의 가능한 그룹  $C$ 를 다음과 같이 할당한다.

$$p(C_k | x_1, \dots, x_n) = \frac{p(C_k)p(\mathbf{x} | C_k)}{p(\mathbf{x})}.$$

독립 일 때 그룹  $C$ 의 조건부 분포는 아래와 같다.

$$p(C_k | x_1, \dots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^n p(x_i | C_k),$$

여기서  $Z = p(\mathbf{x})$ 로  $x_1, \dots, x_n$ 에만 의존하는 규모요소이다. 새로운 입력벡터는 가장 가능성 높은 집단에 속하는 것으로  $C_k$ 에 대해서 다음 식을 통해 최대 확률을 갖는 그룹  $k$ 를 찾아낸다[14].

$$\hat{y} = \operatorname{argmax}_{k \in (1, \dots, K)} p(C_k) \prod_{i=1}^n p(x_i | C_k).$$

### 2.4 서포트벡터머신

(Support vector machine, SVM)

서포트벡터머신은 서포트벡터를 통해 학습자료에 대한 오차를 최소화하여 분류하는 비확률적 이진분류 기계 학습방법이다. 데이터의 사상된 공간상에서 거리를 측정하여 두 그룹 사이의 중심과 최적의 초평면을 구해 두 그룹으로 나누는 법을 학습한다. 최적의 경계면은 양 그룹으로부터 가장 멀리 떨어진 공간이다[15]. 본 연구에서는 R의 'e1071'패키지를 통해 나이브베이즈와 서포트벡터머신을 이용한 분류를 수행하였다[16].

### 2.5 로지스틱회귀모형 (Logistic Regression)

로지스틱회귀모형은 결과가 범주형일 경우 사용되는 기법으로 이항, 다항, 순서 로지스틱회귀모형이 존재한다. 본 연구에서는 이항 로지스틱회귀(binomial logistic

regression)가 사용되었다. 이항 로지스틱 회귀 분석에서 2개의 범주는 0과 1로 나타내어지고 각각의 범주로 분류될 확률의 합은 1이 된다.  $y = w \times x + b$ 의 식에서  $y$ 를 확률에 관한 식으로 만든 다음  $f(x)$ 를 로짓변환을 취하여 입력 값의 범위가  $[-\infty, \infty]$ 일 때, 출력 값의 범위를  $[0, 1]$ 로 조정한다.  $x$ 의 값의 증감과 무관하게  $f(x)$ 는 0과 1 사이의 값을 갖는 확률 값으로 사용할 수 있다. 로지스틱 회귀분석은 돌발홍수가 발생할 경우와 발생하지 않을 경우를 예측하기보다는 그 지역에 돌발홍수가 발생할 확률을 계산해 준다는 장점이 있다[17].

### 2.6 평가방법

본 연구에서는 빅데이터 분석 방법의 예측성능을 평가하기 위해 정확도, kappa계수, TP Rate, FP Rate, F-Measure를사용하였다. 실제값과 예측값의 일치여부를 나타낸 혼동행렬(Confusion Matrix)로부터 위의 값을 도출해낼 수 있다. TP Rate는 정답을 정답으로 분류할 비율이므로 높은 값을 나타낼수록 성능이 좋음을 의미한다. 정확도, kappa계수, F-Measure 또한 값이 클수록 좋은 성능을 나타낸다. 반면, FP Rate는 오답을 정답으로 분류하고 있는 비율이므로 낮은 값을 나타낼수록 성능이 좋음을 의미한다.

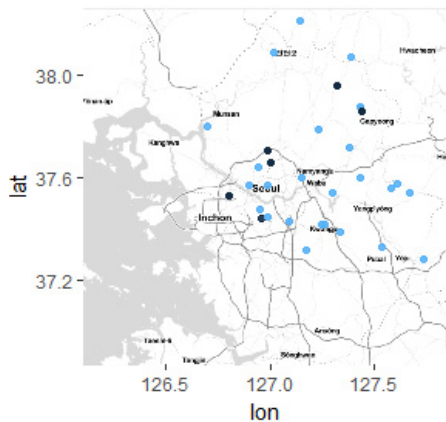
최적 빅데이터 분석 방법 선정을 위해 돌발홍수 발생 선행 6시점에 대해 시점별 재현성과 2009년부터 2012년까지 돌발홍수 실제 예보 횟수와 예측 예보 횟수의 비교하였다.

## 3. 자료 분석

### 3.1 연구자료

본 연구 자료는 2009년부터 2012년까지 국민안전처 국가재난정보센터에 보고된 38건의 홍수 피해 사고 자료를 이용하였다[7]. 돌발홍수 유무를 모의하기 위한 지표 수문모형으로 TOPLOTS가 사용되었으며 돌발홍수 예보를 위한 수문기상정보로 강우량, 지표유출량, 토양수분 상태가 이용되었다.

[Fig. 1]은 서울·경기지역에 4년 동안 발생한 돌발홍수로 인한 사고 위치정보이다. 사고가 1회 발생한 지역은 파랑색, 2회 발생한 지역은 남색으로 표현하였다.



[Fig. 1] The location of flash flood

### 3.2 빅데이터 분석

돌발홍수 발생 선행 6시간의 수문기상정보(강우량, 지표유출량, 토양수분상태)로 입력변수들의 특성을 파악하고 자료의 차원을 축소하기 위해 요인분석을 실시하였다. 요인분석 결과는 <Table 1>이다. 요인분석 결과 3개의 수문기상정보의 선행 6시점인 18개의 변수가 요인 3개로 축소되었다. 요인1은 토양수분 상태, 요인2는 장기요인에 의한 강우량과 지표유출량, 요인3은 단기요인에 의한 강우량과 지표유출량을 의미한다. 요인1은 전체 분산의 약 33%를 요인2는 전체분산의 약 21%를 그리고 요인3은 약 20%를 설명하므로 세 개의 요인으로 전체분산의 약 74.5%가 설명된다. 토양수분 상태는 돌발홍수 시 토양이 흡수할 수 있는 수분량 및 증발산량에 따라 변하는 값으로 지표수문모델에서 가장 많은 분산을 차지하고 있다. 지표유출량과 강우량은 장기요인과 단기요인으로 구분된다.

요인분석을 통해 얻은 세 개의 요인으로 의사결정나무, 랜덤포레스트, 나이브베이즈, 서포트벡터머신, 로지스틱 회귀분석을 실시한 결과는 <Table 2>이다. 의사결정나무는 DT, 랜덤포레스트는 RF, 나이브베이즈는 NB, 서포트벡터머신은 SVM, 로지스틱회귀모형은 Lo 로 표시하였다. 나이브베이즈가 정확도 100%, 그 다음으로 로지스틱회귀모형이 92%, 서포트벡터머신이 87%, 의사결정나무와 랜덤포레스트가 84%의 순으로 예측정확도가 높았다. Kappa계수는 관측값과 예측값의 일치도를 평가하는 계수로 나이브베이즈가 100%, 로지스틱회귀모형이

83%, 서포트벡터 머신이 72% 순으로 높게 나타났다. TP Rate와 FP Rate를 비교해본 결과 랜덤포레스트가 가장 우수하고 그다음이 서포트벡터머신으로 나타났다. F-Measure 값에서는 랜덤포레스트가 가장 값이 높게 나타났다.

<Table 1> The result of factor analysis

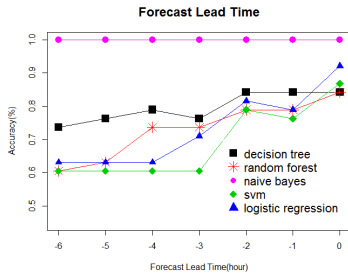
variable	Fac.1	Fac.2	Fac.3
Precipitation -1	-0.095	0.278	0.648
Precipitation -2	-0.081	0.398	0.638
Precipitation -3	-0.069	0.588	0.484
Precipitation -4	-0.061	0.775	0.271
Precipitation -5	-0.055	0.875	0.088
Precipitation -6	-0.051	0.816	0.008
Surface runoff -1	-0.060	0.011	0.806
Surface runoff -2	-0.047	0.090	0.869
Surface runoff -3	-0.040	0.274	0.760
Surface runoff -4	-0.038	0.487	0.558
Surface runoff -5	-0.041	0.644	0.352
Surface runoff -6	-0.049	0.654	0.227
Soil moisture -1	0.994	-0.026	-0.071
Soil moisture -2	0.995	-0.038	-0.076
Soil moisture -3	0.995	-0.054	-0.075
Soil moisture -4	0.994	-0.072	-0.070
Soil moisture -5	0.993	-0.087	-0.063
Soil moisture -6	0.991	-0.097	-0.057
SS loadings	5.965	3.804	3.640
Proportion Variance	0.331	0.211	0.202
Cumulative Variance	0.331	0.543	0.745

<Table 2> The result of performance evaluation

	ACC.	Kappa	TPR.	FPR.	F
DT.	0.842	0.645	1.000	24.000	1.922
RF.	0.842	0.662	5.000	22.000	8.413
NB.	1.000	1.000	1.000	0.000	0.000
SVM	0.868	0.721	3.000	22.000	5.357
Lo.	0.921	0.829	1.000	24.000	1.929

돌발홍수는 짧은 강우 시간과 강한 강우 강도에 대피할 시간이 부족한 것이 특징이다. 따라서 선행 시간별 돌발홍수 발생 여부를 판단하여 재현성을 구현 하는 것이 보다 현실적인 접근법이다. 분석방법에 따른 재현성은 [Fig. 2]이다. 나이브베이즈는 돌발홍수 발생 선행 6시간 모두 정확도가 100%로 과대추정이 의심된다. 돌발홍수 발생 선행 2시간을 기준으로 로지스틱 회귀모형이 82%, 의사결정나무가 84%, 랜덤포레스트와 서포트벡터머신이

79%의 순으로 정확도가 높았다. 나이브베이지를 제외하면 예측 선행시간이 길어질수록 돌발홍수 발생여부 판단 정확도가 낮아지는 경향을 보인다.



[Fig. 2] Repeatability test

돌발홍수와 같은 자연재해 발생여부 판단에서 과소추정은 피해 사고 대비 목적에 어긋나며, 과대추정은 시민들에게 불편을 초래한다. 이에 본 연구의 과소추정과 과대추정 여부 모의를 위해 훈련자료인 38건의 돌발홍수 피해 자료를 포함한 4년간 자료에서 돌발홍수로 경보하는 횟수를 알아보았다<Table 3>. 2009년부터 2012년까지 4년간 비가 내린 날은 273일이다. 모의결과 로지스틱 회귀모형은 54일, 랜덤포레스트는 78일, 서포트벡터머신은 80일, 나이브베이지는 92일, 의사결정나무는 156일로 돌발홍수 발생을 예보 하였다. 의사결정나무는 비가 온 날의 57%를 돌발홍수로 예보했으므로 과대추정임을 알 수 있다.

<Table 3> The number of flash flood forecast time during 4 years

	Rainy days	flash flood Warnings	Warning rate
DT.	273	156	57.1%
RF.	273	78	28.6%
NB.	273	92	33.7%
SVM	273	80	29.3%
Lo.	273	54	19.8%

#### 4. 논의 및 결론

돌발홍수는 강우유출수가 하천으로 모여드는 유역이 좁은 지역에 집중호우로 유입되는 물의 양이 급증하여

나타난다. 돌발홍수는 유속이 빨라 안전사고에 대비할 수 있는 시간이 부족하다. 본 연구에서는 빅데이터 분석 방법을 이용하여 돌발홍수 발생 예보를 위한 분석을 수행하였다.

국가재난정보센터에 보고된 38건의 돌발홍수 피해사례와, 돌발홍수 발생 선행 6시점의 수문지표모형을 연구자료로 이용하였다. 분석방법으로는 의사결정나무, 랜덤포레스트, 나이브베이지, 서포트벡터머신, 로지스틱회귀모형이 고려되었다.

로지스틱회귀모형의 돌발홍수 발생 예보 정확도는 92%이고 재현성분석에서는 선행 2시간 까지 예측 정확도 80%를 보였다. 2009년부터 2012년까지 4년 동안 돌발홍수 예보를 모의한 결과는 돌발홍수 예보 횟수는 54회로 다른 방법에 비해 상대적으로 낮았다.

랜덤포레스트는 돌발홍수 발생 예보 정확도가 87%이고 선행 2시간의 재현성 예측정확도가 82%였으며, 4년 동안 돌발홍수 예보 모의는 78회였다. 로지스틱 회귀모형에 비해 예보 정확도가 낮고 모의 횟수가 높지만 재현성은 로지스틱에 비해 높았다.

나이브베이지는 훈련자료의 돌발홍수 예보 정확도가 100%이고, 선행시간의 재현성도 100%로 로지스틱회귀모형이나 랜덤포레스트에 비해 높았지만, 4년간 모의된 돌발홍수 예보의 수가 92회로 랜덤포레스트와 로지스틱회귀모형에 비해 높으므로 과적합 된 것으로 판단된다.

본 연구에서는 적은 수의 사례에서의 빅데이터 분석 방법의 적용결과를 살펴보았다. 국가재난센터에 매년 돌발홍수로 인한 피해보고사례가 보고되고 있다. 본 연구에서 사용된 자료는 2009년부터 2012년까지 보고된 38건의 돌발홍수 피해사례로 한정되어 분석을 위한 훈련자료와 검증자료 구축에 한계가 있었다. 장기간 자료가 수집되어 구축된다면 더욱 정확한 빅데이터 분석을 수행할 수 있다.

#### REFERENCES

- [1] H. Chang & W. T. Kwon, "Spatial variations of summer precipitation trends in South Korea, 1973-2005," Environmental Research Letter, Vol. 2, No.4, pp.1-9, 2007.
- [2] S. Lee, W. T. Kwon, "A variation of summer

- rainfall in Korea.”, Journal of Korean Geographical Society, Vol.39, No.6, pp.819-832, 2004.
- [3] Korea Meteorological administration, “Special report about abnormal climate” · (<http://www.climate.go.kr/index.html>), 2010.
- [4] D. H. Bae & J. H. Kim, “Development of Korea Flash Flood Guidance System : ( I ) Theory and System Design.”, KSCE Journal of Civil Engineering, Vol.27, No.3B, pp.237-243, 2007.
- [5] J. H. Lee, H. D. Jun, M. J. Park & J. H. Jung, “Flash flood risk assessment using PROMETHEE and Entropy method.” Journal of Korean Society of Hazard Mitigation, Vol. 11, No. 3, pp.151-156, 2011.
- [6] S. Yoon, S. Choi, B. J. Lee & Y. Choi, “Study on Statistical Methods for the Development of Flash Flood Index.”, Journal of Korean Society of Hazard Mitigation, Vol.15, No.6, pp.189-197, 2015.
- [7] B. J. Lee, S. Choi, S. Yoon & Y. Choi, “Evaluation of TOPLATS Land Surface Model Application for Forecasting Flash Flood in mountainous areas.”, Journal of Korea Water Resources Association, Vol.49, No.1, pp.19-28, 2016.
- [8] Korea Meteorological administration, “Meteorological Technology & policy” · (<http://www.climate.go.kr/index.html>), 2009.
- [9] J. Choi, S. T. Han, H. Kang & E. Kim, “Data Mining Decision Tree Analysis Using Answer Tree.” SPSS academy, pp.17-23, 1998.
- [10] M-H. Lee & M-G. Kim, “Meteorological information analysis algorithm based on weight for outdoor activity decision-making”, Journal of Digital Convergence, Vol.14, No.3, pp.209-217, 2016.
- [11] T. Therneau, B. Atkinson & B. Ripley. Package “rpart: Recursive Partitioning and Regression Trees. R package version 4.1 - 10”, 2015.
- [12] C. Park, “A simple diagnostic statistic for determining the size of random forest.”, Journal of the Korean Data & information Science Society, Vol. 27, No.4, pp.855-863, 2016.
- [13] A. Liaw & M. Wiener, “Classification and regression by randomForest.” Vol.2, No.3, pp.18-22, 2002.
- [14] H-S. Seo & S-Y. Lee, “A Model to Infer Users’ Behavior Patterns for Personalized Recommendation Service based Context-Awareness”, Journal of Digital Convergence, Vol.10, No.2, pp.293-297, 2012.
- [15] H. Lee, S-H. Chung, & E-J. Choi, “A Case Study on Machine Learning Applications and Performance Improvement in Learning Algorithm.”, Journal of Digital Convergence, Vol.14, No.2, pp.245-258, 2016.
- [16] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, C. Chang & C. Lin, “Misc Functions of the Department of Statistics. Probability Theory Group (Formerly: E1071)”, TU Wien, 2015.
- [17] B. Chae, W. Kim, C. Cho, K. Kim, C. Lee, & Y. Choi, “Development of a Logistic Regression Model for Probabilistic Prediction of Debris Flow”, The Journal of Engineering Geology, Vol.14, No.2, pp.211-222, 2004.

#### 박 다 인(Park, Da In)



- 2014년 8월 : 대구대학교 전산통계학과 (학사)
- 2016년 9월 ~ 현재 : 대구대학교 통계학과 (석사)
- 관심분야 : 통계, 인공지능
- E-Mail : joibox54321@naver.com

#### 윤 상 후(Yoon, Sang Hoo)



- 2006년 12월 ~ 2007년 12월 : The university of Auckland 인턴연구원
- 2011년 8월 : 전남대학교 통계학과 (이학박사)
- 2012년 12월 ~ 2013년 11월 : The university of Southampton 박사후연구원
- 2013년 12월 ~ 2015년 2월 : 의료방사선안전연구원 전임연구원
- 2015년 2월 ~ 2016년 2월 : 한국외국어대학교 차세대도시농림융합기상사업단 선임연구원
- 2016년 3월 ~ 현재 : 대구대학교 전산통계학과 조교수
- 관심분야 : 시공간모형, 극단치 분포, 통계학습
- E-Mail : statstar@daegu.ac.kr