

공공 데이터의 빅데이터 분석을 통한 사회 안전망 시스템

이선의*, 정준희*, 차경현*, 손기준**, 김상지***, 김진영*

Social Safety Systems through Big Data Analysis of Public Data

Sun Yui Lee*, Jun Hee Jung*, Gyeong Hyeon Cha*, Ki Jun Son**, Sang Ji Kim***, and Jin Young Kim*

요 약

본 논문은 빅 데이터 분석을 이용하여 산악 안전사고를 예방하기 위하여 사고 예측 모델을 제시하였다. 산악 안전사고의 축적된 데이터를 파악하기 쉽게 그래프로 나타내었다. 사고가 발생하는 패턴을 알기 위하여 산악 안전사고 발생 건수의 연도별 분석, 연간 월별 사고 발생 건수, 요일별, 시간대별 분석을 수행하였다. 나타난 그래프를 이용하여 산악 안전사고의 영향을 미치는 변수들을 가중치 모델링을 통하여 사고 예측 모델을 구성하였다. 산악 지역의 사고 다발 구역에 제시한 모델을 적용하여 예측 모델의 성능을 검증하였다.

Key Words : Big Data, Crime Pattern, Secure Sharing

ABSTRACT

This paper proposed an accident prediction model in order to prevent accidents in mountain areas using a big data analysis. Data of accidents in mountain areas are shown as graphs. We have analyzed cases: the number of accidents per year, day of week, time of day to find patterns of the negligent accident in mountain areas. The proposed prediction model consists of weighted variables of the accident in mountain through visualized big data analysis. The model of danger index performance is demonstrated by showing accident-prone areas with weighted variables.

I. 서 론

데이터의 디지털화가 빠르게 진행됨에 따라 거대한 데이터 구조가 구축되어지고 있다. 데이터의 수집과 분류, 분석에 따른 이 데이터들의 최적화에 따른 각각의 정보들의 가치가 더 세밀해질 가능성을 제시하고 있다. 그 데이터들은 각각뿐만 아니라 또 다른 데이터 수요와 결합하여 새로운 정보를 제공할 가능성을 내포하고 있다 [1-3].

지금까지의 데이터 클라우드 저장은 텍스트 또는 수동적인 암호화된 코드들로 구성되어져 왔다. 이러한 데이터들은 분석하기가 복잡하고 특정 수요를 만족하지 못하면 활용할 가능성을 쉽게 찾을 수 없다.

하지만 빅 데이터 플랫폼을 이용하여 이 데이터를 분류 정제하여 쉽게 예상할 수 있는 정보로 활용할 수 있다. 공공의 이익을 위하여 수집되어 축적된 데이터를 예로 들면 특정 지역의 최근까지 수집된 Crime Pattern(범죄 발생 통계)를

바탕으로 범죄 발생 확률을 예측하여 순찰차의 순찰 경로를 지정하여 범죄 예방에 활용할 수 있다. 또한 Secure Sharing 및 특정 범죄가 자주 발생하는 통계를 근거로 원인을 추적하여 개선할 수 있게 쓰일 수 있다. 강도 및 성범죄의 시간별 발생 통계를 바탕으로 그 지역의 수은등을 LED 램프로 교체하여 실제 범죄 발생 통계의 변화를 바탕으로 범죄 발생 예방을 검증할 수도 있다 [4-6].

이와 같이 빅 데이터 분석을 통하여 공공의 이익을 위한 정보 취득과 분석에 쉽게 활용할 수 있는 방법이 필요하다. 본 논문에서는 공공에게 공개된 전과정보를 통하여 빅 데이터 분석을 수행하여 안전사고 예방에 쓰일 수 있는 방법을 제시한다. 각 기지국은 유저들의 서비스 활용 종류를 파악하고 있고 서비스 지역 내의 대략적인 위치 또한 파악이 가능하다. 따라서 국가 기관이 파악한 사고 통계 자료와 기지국이 가지고 있는 전과 활용 정보와 결합하여 안전사고 예방에 활용할 수 있다. 이처럼 국내뿐만 아니라 국외에서도 빅 데이

*이 논문은 2014년 미래창조과학부의 재원으로 SW융합기술고도화 사업의 지원을 받아 수행된 연구임(S0170-15-1081).

*광운대학교 전파공학과 소속 유비쿼터스 통신 연구실 (sunyui22@naver.com, zmsmsm@naver.com, bizkit-@nate.com, jinyoung@kw.ac.kr)

** (주)더아이엠씨 (kjsn@theimc.co.kr)

*** (주)위니텍 (chocoajong@gmail.com, goldie64@naver.com)

접수일자 : 2015년 11월 16일, 최종게재확정일자 : 2015년 12월 24일

터 분석을 이용하여 다양한 사회 안전망 구축을 위한 연구가 활발히 이루어지고 있다.

이에 따라 본 논문에서는 국가재난정보센터에서 제공하는 재난 통계 및 안전사고 자료를 바탕으로 전과 정보 데이터와 결합한 실시간 사고 예측 시스템을 제안한다. 2장에서는 산악지역에서의 안전사고 데이터를 분석한다. 분석 데이터를 바탕으로 주요 연령, 사고 발생 시간, 지역에 따른 사고 유형에 따른 전과 정보와 결합한 사고 예방 알고리즘을 제시한다. 3장에서는 사고 예측을 위한 알고리즘을 활용하여 공공 데이터 정보와 분석을 통한 데이터의 쉬운 분석을 위한 시각화 서비스 모델을 제시한다. 4장에서는 전과 정보를 활용하여 사용자의 데이터 서비스 활용에 따른 사용자 위치, 시간에 따른 통계 자료와 결합한 예측의 오차 보정 시스템을 제시한다.

II. 산악 안전사고 빅 데이터 정리

사회 안전망 구축을 위한 빅 데이터 분석을 위하여 안전 사고가 많이 일어나고 있는 산악지역의 사고 통계를 분석하고자 한다. 등산인구의 증가로 국립공원에 방문객은 해마다 가파르게 증가하고 있다. 이와 함께 안전 수칙을 지키지 않아서 생기는 사고와 노령의 인구가 맞지 않는 코스를 선택하여 벌어지는 다양한 산악지역의 사고가 증가하고 있다. 따라서 이와 같은 사고의 예방 및 긴급구조 활동의 신속한 이동을 위하여 축적된 통계 데이터를 바탕으로 사고를 예측한다.

다양한 공공 데이터가 축적되어 가고 있는 현대 시대에는 각 나라마다 공공 데이터를 개방하여 다양한 서비스를 수행할 수 있게 제공하고 있다. 본 논문에서는 대한민국에서 제공하는 공공데이터 중 등산객 현황을 분석하여 빅 데이터 분석을 통한 안전사고 예방에 활용한다. 범죄 예방에 사용되는 빅 데이터 분석은 범죄 종류, 지역, 시간별, 발생 사례를 토대로 지역을 시간대별로 차별화된 순찰 루트를 결정하여 사고를 예방하는 방식으로 이용된다. 이와 같이 산악 지역에서의 발생하는 안전사고 데이터를 분석하여 안전사고를 예방한다.

본 장에서는 산악 지역의 사고 발생 데이터를 바탕으로 안전사고 발생을 예측하는데 활용한다. 그림 1은 년도별 산악사고 발생인원을 나타낸 그래프이다. 2001년부터 2014년까지 등산객의 사고건수가 지속적으로 증가해온 것을 알 수 있다. 보도 자료 및 아웃도어 시장이 지속적으로 성장한 통계를 보아도 등산인구가 가파르게 증가하고 있다는 것을 알 수 있고 등산인구 증가가 적은 경우에는 기상의 영향을 받아 등산 가능 시간이 감소하여 발생한 것으로 보인다. 국가통계포털에서 제공하는 정보를 보면 국립공원 관리공단에서 파악한 전체 등산객 수는 그림 2와 같다. 국립공원의 탐방객 수는 2008년이후로 가파르게 증가하여 2013년에 최대치를 갱신하였고 현재까지 계속 증가하고 있다. 증가하는 등산 인구

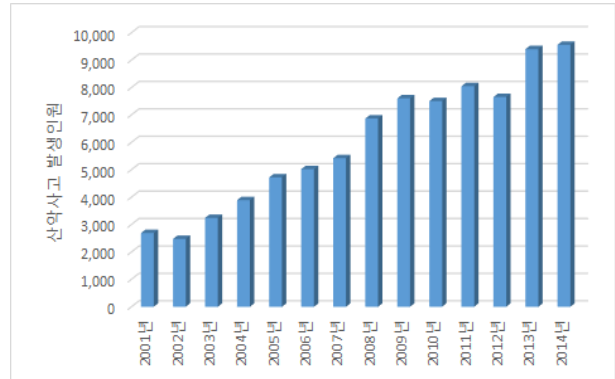


그림 1. 2001~2014년간 산악사고 발생 인원. (단위 : 명)
 <출처 : 국가통계포털>

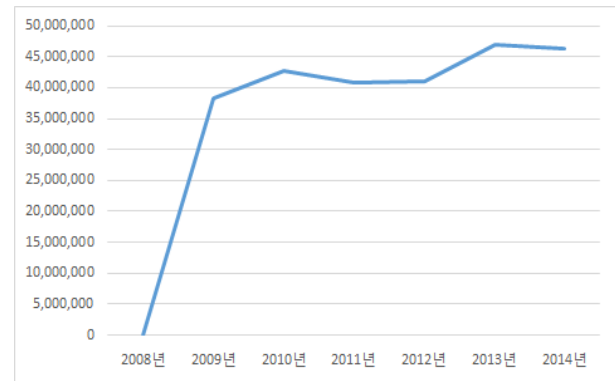


그림 2. 2008~2014년간 국립공원 탐방객 현황. (단위 : 천명)
 <출처 : 국내통계포털 >

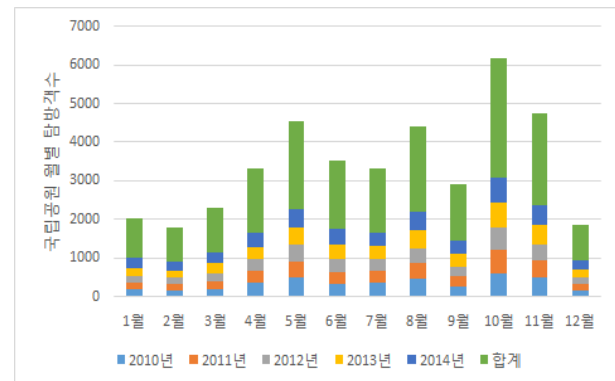


그림 3. 월별 국립공원 탐방객수 현황. (단위 : 만명)
 <출처 : 국립공원관리공단>

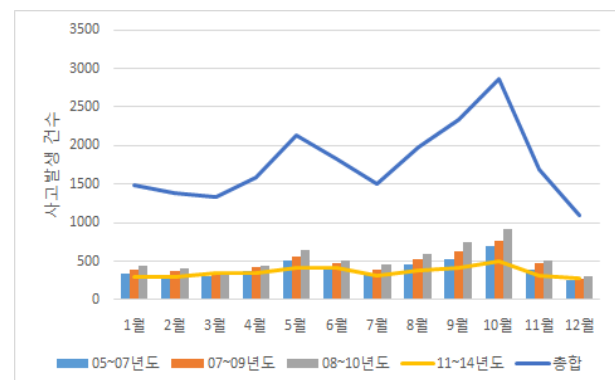


그림 4. 월별 산악 안전사고발생 현황. (단위 : 건수)
 <출처 : 소방방재청, 국가통계포털>

로 인하여 국립공원은 계속 확장 중이고 시설을 확충하고 있는 시점에서 다양한 안전사고가 발생하였고 그 데이터는 시간에 따라 일정한 패턴을 나타내는 것을 알 수 있다. 그림 3을 보면 월별로 국립공원을 방문한 탐방객의 수는 10월 11월 5월 8월 순으로 많은 것을 알 수 있다. 이는 국립공원을 이용하는 등산객의 숫자는 날씨와 계절에 많은 영향을 미치고 있는 것을 알 수 있다. 이에 대한 사고 자료를 분석하여서 그래프로 나타낸 것이 그림 4이다.

연간 분석한 산악 안전사고 자료 등을 바탕으로 그래프로 분석하여 보면 등산객이 많을수록 안전사고가 급증하는 것을 명확하게 알 수 있다. 이 분석 결과 또한 10월에 안전사고가 가장 많이 일어나는 것을 알 수 있고 9월, 5월 순으로 사고가 많이 발생 하였다. 또한 요일별로 발생한 산악 안전사고 데이터를 종합하여 분석한 결과 사고는 주로 등산객이 많이 찾는 주말에 집중적으로 발생하는 것을 알 수 있다.

그림 5는 2008년에서 2010년 3년 동안 발생한 산악 안전사고를 요일별로 분석하여 나타낸 그래프이다. 또한 안전사고 분석을 위하여 샘플로 9월을 기준으로 잡고 요일별로 발생한 사고의 발생건수를 꺾은 선형 그래프로 나타낸 것이다. 이 또한 다년간 분석한 요일별 사고 발생 건수와 일치하는 것을 알 수 있다.

산악 안전사고 발생 시간을 조사하기 위하여 공개된 통계 자료를 바탕으로 시간대별로 사고발생 건수를 조사하였다. 3년 단위로 조사하여 하루 동안 사고가 발생하는 시간을 조사한 데이터를 바탕으로 나타낸 그래프가 그림 6이다. 2006~2008년, 2007~2009년, 2008~2010년 동안의 산악 지형의 국립공원에서 발생한 안전사고를 시간에 따라 평균 건수로 나타낸 것이다. 사고가 주로 발생하는 시간은 빅 데이터 분석을 위하여 중요한 근거가 되기 때문에 1시간 단위로 나타내었다. 주로 등산객들이 주말 산행을 이용할 때 10시부터 등산객이 많이 찾기 때문에 이 시간부터 사고 발생 건수가 증가하기 시작하여 14시에 최고점을 찍고 해가 지기 전에 돌아가기 때문에 18시 이후로 사고 발생이 급격히 감소한다. 이와 같이 등산객이 많이 찾는 계절별, 월별, 요일별, 시간별로 사고를 분석하여 가장 많이 산악 안전사고가 발생하는 지점을 특정 지을 수 있다.

다음으로 산을 많이 찾는 등산객의 산악 안전사고 발생 연령별 분석을 살펴보면 40대와 50대가 가장 사고 발생자가 많은 것을 알 수 있다. 이 통계는 2006~2008년, 2007~2009년 기간 동안 연령별 사고 발생자의 평균을 막대그래프로 나타낸 것으로 그림 7이다. 또한 그림 7의 꺾은 선형 그래프는 2009년 9월에 발생한 사고 연령별 산악 안전사고 발생자수를 나타낸 것이다. 3년간의 평균 안전사고 발생자수와 같이 40에서 50대 연령층에 사고 발생이 두드러지게 나타난 것을 알 수 있다. 2010~2014년 동안 최근 발생한 산악 안전사고 발생 원인을 분석하여 얻은 데이터를 근거로 빅 데이터화하기 위하여 그래프로 나타내면 그림 8과 같다.

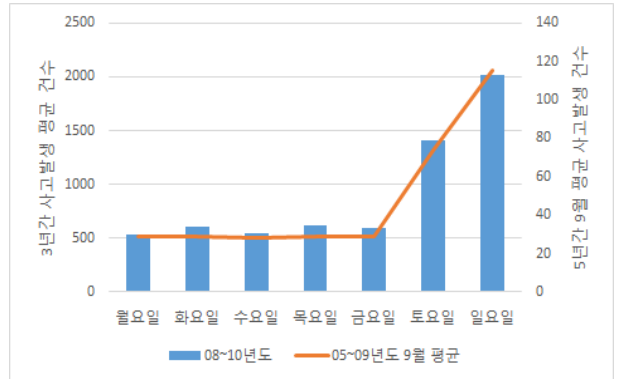


그림 5. 평균 요일별 산악 안전사고 발생현황. (단위 : 건수)
 <출처 : 재난관리정보 DB센터>

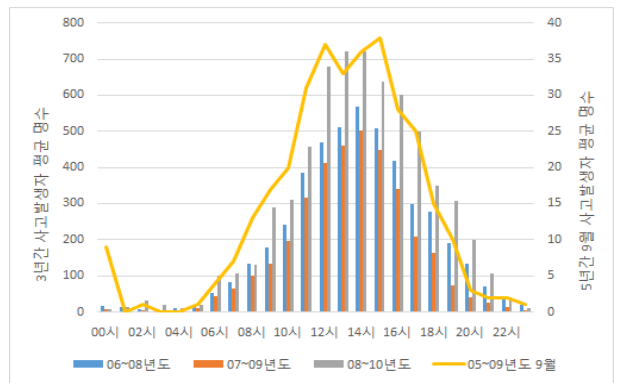


그림 6. 시간대별 산악 안전사고발생 현황. (단위 : 명수)
 <출처 : 소방방재청>

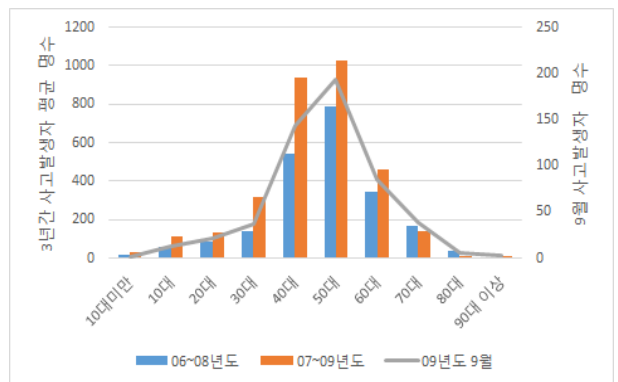


그림 7. 연령별 산악 안전사고발생 현황. (단위 : 명수)
 <출처 : 119 구조 구급 활동 상황 통계>

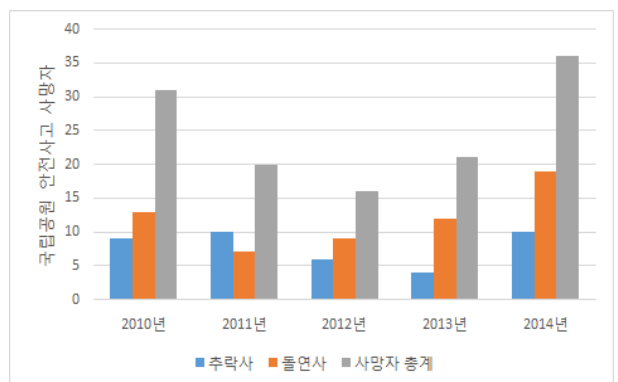


그림 8. 국립공원 안전사고 사망자 현황. (단위 : 명)
 <출처 : 국립공원관리공단>

주요 안전사고 발생 원인은 심장마비 등의 돌연사가 제일 많았고 추락에 의한 추락사가 그 다음 순위이다.

또한 2009~2013년 동안 발생한 산악사고 발생 원인을 부상별로 조사해본 결과 부상자의 경우 골절·상처 1,151명(70.6%), 탈진·경련·저체온 479명(29.4%) 순임을 알 수 있다. 사고로 인하여 구조대의 진입이 어려운 지형에서는 산악 지형에 경우 구조 헬기를 이용한 부상자 이송이 필수적이다. 40~50대 연령의 경우 산악 안전사고 발생 시 부상자의 경우 소방대원의 인력을 통한 이송이 어렵기 때문에 소방헬기 출동이 요청되는 경우가 많다. 그림 9는 산악 안전사고에 따른 전국 소방항공대 활동을 출동 건수에 따라 연도별로 나타낸 그래프이다. 2014년까지 소방헬기의 출동이 계속 증가해 온 것을 알 수 있다. 이는 소방헬기가 착륙할 수 있는 지점이 사고자 이송에 핵심이기 때문에 근처 착륙지점의 유무가 빅 데이터 분석에 필요한 것을 알 수 있다.

Ⅲ. 등산객 안전사고 위험지수

빅 데이터 분석을 위하여 공공 데이터를 수집하여 얻은 자료를 바탕으로 등산객 사고는 2008년부터 등산객이 증가하는 시점부터 매우 급격하게 증가해 온 것을 알 수 있다. 축적된 안전사고 데이터를 분석한 결과 연간 사고 발생 건수는 꾸준히 증가하고 있으며 월별로는 10월, 5월 9월 순으로 사고가 많이 일어났다. 요일로는 평일보다는 주말에 일요일, 토요일 순으로 사고가 전체 사고 발생 건수의 절반이 넘는 53%의 사고가 발생하였다. 시간대별로는 등산객이 많은 11~17시 사이에 전체 발생 건수의 68%를 차지하였다. 이와 같은 발생 시간별 데이터와 사고 다발 지역의 데이터를 같이 분석하여 사고 예방을 위한 위험지수를 산출한다.

국립공원의 무선 단말기의 서비스 내역을 분석하여 사고 발생 예측에 활용한다. 무선 단말기가 주요 사고 다발 지역 기지국에서 신호를 송수신할 경우 위험지수가 올라가고 사용자의 단말이 가지는 연령 데이터를 통한 등반 시간을 계산하여 안전사고를 예방할 수 있게 하는 시스템을 제안한다.

이는 각 산악 안전사고 발생이 많이 일어나는 공간과 시간에 가중치를 더 높게 잡아 위험치 계산을 하고 단말기가 가지는 개인정보와 국립공원 내부의 무선 기지국의 정보를 결합하여 실시간으로 위험도를 예측하여 사용자에게 경고하는 실시간 맞춤 위험지수 계산 모델이다.

표 1은 위험지수를 나타내기 위한 빅 데이터 분석을 통하여 등산객의 사고 발생 예측에 필요한 데이터를 각각 가중치를 갖는 변수로 나타낸 것이다.

식 1은 시간과 지역에 따른 변수를 이용하여 나타낸 결과 값 Ens를 구현식이다. 최종적으로 구한 위험지수 값은 지역에 위치한 기지국의 숫자, 수신 받는 GPS신호의 숫자, 국립공원별 119 출동 시간 등을 나타낸다.

$$Ens = A \times B \times 10 \times C \times T \times Y \times F. \quad (1)$$

나머지 데이터 등은 등산객의 숫자에 영향을 미치는 기상과 기지국의 내의 연령별 가중치 변수이다. 이를 통하여 모델링한 수식 2는 제일 위험군의 위험지수 값을 80~100으로 기준으로 사용하였다.

$$\text{위험지수} = \left[\left(\frac{Ens \times G^2 \times (H-1) \times P}{15 \times (R \times Y)^4} \right) - 10 \right] \div 12. \quad (2)$$

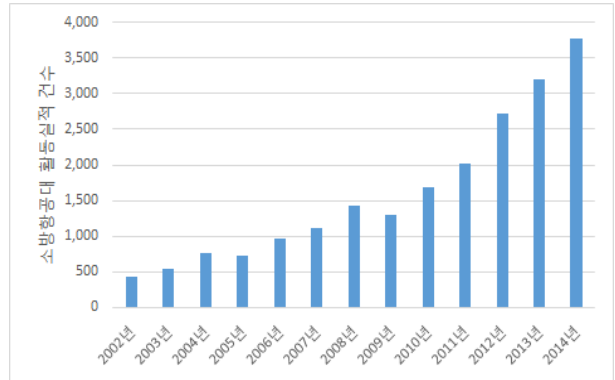


그림 9. 연간 소방헬기 산악구조 출동 건수.
<출처 : 소방방재청, 국가통계포털>

표 1. 산악 안전사고 예측 모델 가중치 변수

구분	가중치
기지국 내 연령수별 : A	<ul style="list-style-type: none"> 40, 50대 : 10 60, 70대 : 7 30, 20, 10대 : 5 기타 : 2
월별 : B	<ul style="list-style-type: none"> 10월 : 10 5월 : 7 9월 : 6 기타 : 2
요일별 : C	<ul style="list-style-type: none"> 일요일 : 10 토요일 : 8 기타 : 2
시간별 : T	<ul style="list-style-type: none"> 11~14시 : 10 14~16시 : 7 10~11시 : 6 16~18시 : 5 8~10시 : 4 기타 : 2
지역별 : D	<ul style="list-style-type: none"> 사고 다발 지역 : 5 일반 조난 지역 : 3 실족 추락 지역 : 2
기지국 수 : R	<ul style="list-style-type: none"> 3개 이상 : 2 3개 이하 : 5
GPS 위성 수 : Y	<ul style="list-style-type: none"> 4개 이상 : 2 4개 이하 : 5
국립공원별 : F	<ul style="list-style-type: none"> 설악산, 한라산 : 5 북한산, 지리산 : 3 기타 : 1
기상 : G	<ul style="list-style-type: none"> 맑음 : 5 흐림 : 3 비 : 2
119 출동 시간 : H	<ul style="list-style-type: none"> 5분 이상 : 10 5분 이하 : 5
소방헬기 출동 시간 : P	<ul style="list-style-type: none"> 10분 이상 : 10 10분 이하 : 5

IV. 산악 지역 위험지수 검정

본 논문에서 제시한 사고 예측 모델을 검정하기 위하여 시뮬레이션한 결과는 그림 10이다. 사고 예측 모델을 사용하여 KBS에서 제공한 산악 지역의 사고 다발 지역 지도를 이용하였다. 지역을 500m 간격으로 분석하였고 지도의 등고선의 단위는 100m이다. 나누어진 지역의 119 출동 시간 및 기지국에서 제공하는 전파정보 데이터와 GPS위성 신호 수신 숫자를 가지고 위험과 경계, 주의, 관심으로 분류하였다. 본 검정에 사용한 데이터는 북한산을 예로 분석한 자료로 앞에서 분석한 시간별 사고 발생 건수를 통하여 사용한 가정은 9월1일부터 10월 10일까지로 일요일, 14시, 기상 맑음, 연령은 40대로 시뮬레이션 데이터를 사용 하였다.

분석한 빅 데이터를 이용하여 파악된 사고 다발지역의 지도를 이용하여 시간에 따른 사고 발생을 예측하여 위험지수로 나타내어 등산객들에게 경각심을 줄 수 있고 사고 발생 시 응급의료 및 구조대에서는 인력 배치를 효율적으로 하여 사고 대응력을 높일 수 있다.



그림 10. 북한산 산악 지역 위험지수 검정 시뮬레이션 결과.
 <출처 : KBS 전국 산악사고 2013>

V. 결론

산악 인구가 지속적으로 늘어나면서 다양한 산악 안전사고가 발생하고 있는 시대이다. 본 논문에서는 축적된 데이터를 이용하여 충분한 분석을 통하여 사고 예방에 사용할 수 있는 빅 데이터 기술을 이용해 새로운 사고 예측 모델을 제시하였다. 산악 사고 사례 및 현황을 분석하여 다양한 그래프로 나타내었고 시간대별로 연별, 월별, 요일별, 시간별로 세분화하여 사고 예측의 정밀성을 높였다. 개인의 생체적 특

성에 따라 사고 발생 확률이 달라질 수 있기 때문에 입장객의 연령을 통신 기지국의 전파정보를 활용하여 위험지수로 나타낸 예측 모델을 통하여 산악 지역 위험지수를 검정하였다. 사고 다발 지역의 더 많은 축적된 데이터와 사고 지역의 기지국 현황에 대한 자세한 파악이 된다면 상세한 사고 발생의 특징을 모델화하여 향후 안전사고 예방에 기여할 수 있을 것이다.

참고 문헌

- [1] V. Luukkala and I. Niemelä, "Enhancing a smart space with answer set programming," in *Semantic Web Rules*, M. Dean, J. Hall, A. Rotolo, and S. Tabet, Eds. Berlin, Germany: Springer-Verlag, 2010, pp. 89103.
- [2] R. Feldman, "Techniques and applications for sentiment analysis," *Commun. ACM*, vol. 56, no. 4, pp. 8289, 2013.
- [3] T. Aihkisalo and T. Paaso, "Latencies of service invocation and processing of the REST and SOAP Web service interfaces," in *Proc. IEEE 8th World Congr. Services*, Jun. 2012, pp. 100107.
- [4] G. Mulligan and D. Gracanin, "A comparison of SOAP and REST implementations of a service based interaction independence middleware framework," in *Proc. Winter Simulation Conf., Austin, TX, USA, Dec. 2009*, pp. 14231431.
- [5] J. Delgado, "Service interoperability in the Internet of Things," in *Internet of Things and Inter-Cooperative Computational Technologies for Collective Intelligence (Studies in Computational Intelligence)*, vol. 460. Berlin, Germany: Springer-Verlag, 2013, pp. 5187.
- [6] L. Ramaswamy, V. Lawson, and S.V. Gogineni, "Towards a quality-centric big data architecture for federated sensor services," in *Proc. IEEE Int. Congr. Big Data, Santa Clara, CA, USA, Jun./Jul. 2013*, pp. 8693.

저자

이 선 의(Sun Yui Lee)

준회원



- 2013년 2월 : 광운대학교 전파공학과 졸업
- 2013년 2월 ~ 현재 : 광운대학교 전파공학과 석박사통합과정

<관심분야> : 가시광 통신, 협력통신, 인지무선통신, 양자통신

정 준 희 (Jun Hee Jung)

준회원



- 2015년 2월 : 광운대학교 전자융합 공학과 졸업
- 2015년 3월 ~ 현재 : 광운대학교 전과 공학과 석사과정

<관심분야> : LBS, 협력통신, 5G 통신, 인지무선통신, 빅 데이터, 무선에너지 하베스팅

차 경 현(Gyeong Hyeon Cha)

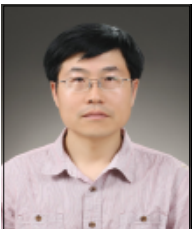
준회원



- 2014년 7월 : 광운대학교 전자융합 공학과 졸업
- 2014년 8월 ~ 현재 : 광운대학교 전과 공학과 석박사통합과정

<관심분야> : 데이터마이닝, 디지털통신, 5G 이동통신

손 기 준(Ki Jun Son)



- 2005년 2월 : 경북대학교 컴퓨터공학과 공학박사 수료
- 2011년 10월 : 에이투텍 플라톤 개발그룹 차장
- 2013년 8월 ~ 현재 : (주)더아이엠씨 빅데이터팀 부장

<관심분야> : 자연어처리, 빅데이터 수집 및 분석

김 상 지(Sang Ji Kim)



- 1990년 2월 : 동서대학교 소프트웨어 공학과 석사
- 2006년 8월 : 쓰리비시스템 대리
- 2009년 6월 ~ 현재 : (주)위니텍 해외사업부 과장

<관심분야> : 신호처리, 영상처리, 패턴인식, 암호화, Water Mark

김 진 영(Jin Young Kim)

종신회원



- 1998년 2월 : 서울대학교 전자공학과 공학박사
- 2001년 2월 : SK텔레콤 네트워크연구소 책임연구원
- 2001년 3월 ~ 현재 : 광운대학교 전자융합공학과 교수

<관심분야> : 디지털통신, 가시광통신, UWB, 부호화, 인지무선통신, 4G 이동통신