

날씨 변화에 따라 교통사고 예방을 위한 예측모델에 관한 연구

정영석^{1*}, 박구락¹, 김진묵²

¹공주대학교 컴퓨터공학과

²선문대학교 IT교육학부

Study on predictive modeling of incidence of traffic accidents caused by weather conditions

Young-Suk Chung¹, Rack-Koo Park¹, Jin-Mook Kim²

¹Division of Computer Science and Engineering, Kongju National University

²Division of IT Education, Sunmoon University

요약 교통사고는 다양한 요인으로 인해 발생한다. 그 중에는 교통사고가 발생할 당시의 기상상태가 있다. 기상 상태에 따라 교통사고로 인해 발생하는 사망자의 비율은 차이가 있다. 교통사고로 인한 사망자의 수를 줄이려면 기상 상태에 따라 발생될 교통사고 발생 수를 예측 하는 것이 필요하다. 본 논문은 기상 상태에 따른 교통사고 발생 빈도수를 예측하는 모델링을 제안한다. 예측 모델링의 이론으로는 마코프 프로세스를 적용하였다. 제안된 모델링에 실제 데이터를 적용하여 교통사고 발생 수를 예측 하였고, 실제 발생 수와 비교하였다. 본 논문은 기상 변화에 따른 교통사고 정책수립에 도움을 줄 것이다.

주제어 : 교통사고, 예측모델, 마코프 프로세스, 날씨 변화

Abstract Traffic accidents are caused by a variety of factors. Among the factors that cause traffic accidents are weather conditions at the time. There is a difference in the percentage of deaths according to traffic accidents, due to the weather conditions. In order to reduce the number of deaths due to traffic accidents, to predict the incidence of traffic accidents that occur in response to weather conditions is required. In this paper, it propose a model to predict the incidence of traffic accidents caused by weather conditions. Predictive modeling was applied to the theory of Markov processes. By applying the actual data for the proposed model, to predict the incidence of traffic accidents, it was compared with the number of occurrences in practice. In this paper, it is to support the development of traffic accident policy with the change of weather.

• **Key Words** : Traffic accidents; predictive models; Markov process; weather condition,

1. 서론

교통사고 중 기상 상태가 교통사고 사상자에 미치는 영향은 크다. 도로교통공단에서 발간하는 교통사고 요인

분석(2013년판)을 보면, 사망사고를 기상 상태 별 발생건 수 구성비를 비교해보면 맑은 날 보다는 흐린 날, 비 오는 날, 안개 낀 날에 상대적으로 많이 발생한 것으로 조

*교신저자 : 박구락(ecgrpark@kongju.ac.kr)

접수일 : 2014년 1월 10일, 수정일 2014년 2월 13일, 게재확정일 : 2014년 2월 16일

사되었다. 다음의 표 1은 기상 상태 별 교통사고 통계자료이다[1].

[Table 1] Traffic accidents by weather conditions

기상상태	구분	발생건수			사망자수		부상자수	
		(건)	구성비	상대사망 사고율	(명)	구성비	(명)	구성비
계		5,165	100	1.00	5,392	100	3,273	100
맑음		4,067	78.7	0.94	4,233	78.5	2,545	77.8
흐림		392	7.6	1.56	407	7.5	253	7.7
비		587	11.4	1.18	625	11.6	420	12.8
안개		31	0.6	4.25	35	0.6	19	0.6
눈		61	1.2	0.95	63	1.2	33	1.0
기타/불명		27	0.5	0.83	29	0.5	3	0.1

현재 기상 상태와 교통사고의 관련성에 관한 연구는 이루어지고 있으나, 예측에 대한 연구는 부족하다. 교통사고 사상자의 수를 줄이려면 기상 상태에 따라 발생할 교통사고발생 수를 예측 하는 것이 필요하다. 본 논문에서는 기상 상태에 따라 교통사고 발생 빈도수를 예측하는 모델을 제안하였다. 그리고 실제 데이터를 적용하여 교통사고 발생 수를 예측하였고, 실제 발생 수와 비교하였다. 예측 모델링의 이론으로 마코프 프로세스를 적용하였다. 본 논문의 구성은 다음과 같다. 2장에서 관련연구에 대해 논의한다. 3장에서 기상상태에 따른 교통사고 빈도수 예측 모델링을 제안한다. 4장에서 실제 데이터를 적용하여 실험한다. 마지막으로 결론 및 향후 연구과제에 대해 논의 한다.

2. 관련연구

2.1 기상 상태와 교통사고의 관련성에 대한 연구

기상 상태와 교통사고의 관련성에 대한 연구는 다음과 같다. 악천후 시 발생하는 교통사고 원인을 분석하기 위해 일반국도에서 발생하는 사고자료를 바탕으로 악천후 시 기상 상태 별 교통사고를 분석하여 도로안전시설 현황 및 문제점을 파악한 연구가 진행되었다[2]. 캐나다 오타와에서 발생한 교통사고와 기상상태의 데이터를 분석하여, 기상 상태에 따라 발생하는 교통사고 중 충돌사고의 위험도에 대한 연구가 있다[3]. 기상 데이터와 교통사고 발생 데이터를 이용하여 비가 오는 날 발생한 충돌사고 위험성을 조사한 연구가 있다[4]. 그 외에도 다양한 기상 상태와 교통사고의 연관성에 관한 연구가 있으나,

주로 기상상태에 따른 교통사고를 분석하는 연구가 진행되었고, 기상 상태에 따른 교통사고를 예측하는 연구는 부족하다.

2.2 마코프 프로세스

기상상태에 따른 교통사고 발생을 예측 하는 이론으로 마코프 프로세스를 적용하였다. 마코프 프로세스는 미래를 예측하는 다양한 연구에 적용되고 있다. 마코프 프로세스는 상태간의 전이가 이전 n 개의 상태에 의존하는 프로세스를 의미한다. 마코프 프로세스를 $X(t)$ 라 하면 임의의 시간 $t_1 < t_2 < \dots < t_k < t_{k+1}$ 에 대해 $X(t)$ 가 이산 값을 의미 할 때를 마코프 체인이라 한다. 수식 (1)은 마코프 프로세스가 이산값 일 때를 표현한 것이다 [5,6,7].

$$P[a < X(t_{k+1}) = x_{k+1} | X(t_k) = x_k, \dots, X(t_1) = x_1] \quad (1) \\ = P[X(t_{k+1}) = x_{k+1} | X(t_k) = x_k]$$

t_k 는 현재, t_{k+1} 은 미래, t_1, \dots, t_{k-1} 은 과거의 시점을 의미한다. 마코프 프로세스는 다양한 예측 연구에 활용되고 있다.

3. 기상 상태에 따른 교통사고 빈도수 예측 모델링

본 논문은 다음의 두 가지 요인을 적용 하여 기상 상태를 적용한 교통사고 예측 모델링을 구성하였다.

첫째: 기상상태에 따라 발생하는 교통사고 분류

둘째: 예측 모델링의 이론으로 마코프 프로세스 적용

두 가지 요인을 적용하여 기상 상태를 적용한 교통사고 예측 모델링은 다음의 그림 2과 같다.

각 단계는 다음의 순서로 되어있다.

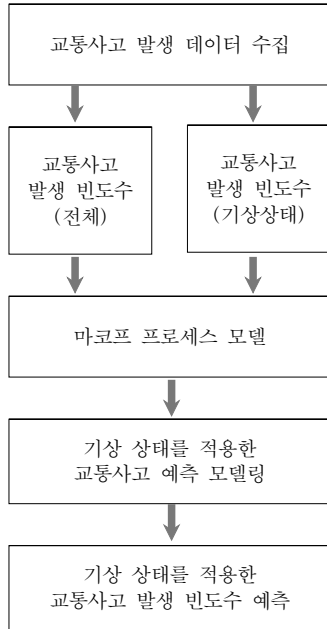
첫 번째 : 교통사고 발생 데이터를 수집한다.

두 번째 : 교통사고 발생 데이터를 다음과 같이 나눈다.

- ④ 매년 발생한 전체 교통사고 빈도수
- ⑤ 매년 각 기상 상태(예: 맑음, 흐림, 눈, 비 등)에 따라 발생한 교통사고 발생 수

세 번째 : 각 교통사고 발생 수를 마코프 프로세스에 적용한다.

마코프 프로세스는 임계값, 상태들의 집합인 상태집합, 초기확률, 전이행렬로 구성되어있다.



[Fig. 1] Predictive modeling of the frequency of occurrence of traffic accidents to which is applied a weather condition

- 상태 집합 : 상태는 교통사고 데이터 들의 빈도수 범위를 의미한다. 통계 자료에 따라 적절한 임계 값을 설정한다. 설정된 임계 값을 바탕으로 상태들을 정의한다. 본 논문은 상태를 두 가지로 정의하였다.
- 초기확률 : 초기확률은 교통사고 예측 모델링 전체 상태 중, 초기 상태가 가질 수 있는 확률이다. 최근에 발생한 상태를 초기 상태로 이용한다. 식(2)로 정의 한다.

$$P(S_1, S_2, \dots, S_n) = P\left(\frac{a}{F}, \frac{b}{F}, \dots, \frac{c}{F}\right) \quad (2)$$

a, b, c : 각 상태(S_1, S_2, \dots, S_n)의 데이터들의 발생 횟수

F : a, b, c 의 합. 초기 확률의 총 합은 1이 되어야하는 조건(3)을 만족한다.

$$\sum_{i=1}^N P(S_i) = 1 \quad (S \text{는 상태}) \quad (3)$$

전이 행렬 : 상태 사이의 전이 상태를 확률로 나타낸 것이다. 시간에 따라 데이터들을 나열하고, 상태 집합에 정의된 상태들과 매핑하여 나열 한다. 나열된 하나의 상태가 다른 상태로의 전이 횟수를 구한다. 그리고 이것을 전이 행렬로 나타낸다.

- 열: 하나의 상태에서 다른 상태로의 확률을 의미.
- 행: 행은 확률 값이다. 그래서 합은 1

식(4)의 P 는 전이 행렬로 다음의 조건식(5)를 만족한다.

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} & \dots & P_{1n} \\ P_{21} & P_{22} & P_{23} & \dots & P_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & P_{n3} & \dots & P_{nn} \end{pmatrix} \quad (4)$$

$$\sum_{j=1}^n P_{1j} = 1, \sum_{j=1}^n P_{2j} = 1, \dots, \sum_{j=1}^n P_{nj} = 1, P_{ij} \geq 0$$

$$\sum_{j=1}^n P_{ij} = 1, i = 1, 2, \dots, n \quad (5)$$

교통사고 발생 예측 모델링의 예측 확률은 마코프 프로세스를 적용한다. 마코프 프로세스 식(6)은 다음과 같다[5].

$$P(S_k) = \sum_{i=1}^n P(S_i) P_{ik} \quad (6)$$

$P(S_i)$: 초기 확률 P_{ik} : 전이행렬

본 연구에서는 두 종류의 예측 확률을 계산한다.

- ④ 매년 발생한 전체 교통사고 빈도수를 적용한 교통사고 예측 확률
- ⑤ 매년 각 기상 상태(예: 맑음, 흐림, 눈, 비 등)에 따라 발생한 교통사고 빈도수를 적용한 교통사고 예측확률

네 번째 : 기상 상태를 적용한 교통사고 예측 모델링은 다음과 같다.

본 논문은 전체 상태 예측 확률 값과 기상 상태에 따른 예측 확률 값을 동등하다고 가정하였다.

세 번째 단계에서 계산한 ㉠확률 값에 ㉡의 결과인 각 기상 상태에 따라 예측된 확률 값을 기상 상태를 적용한 교통사고 예측 모델링에 적용한다. 수식(7)은 각 기상상태 별 교통사고 발생 확률이다.

$$\frac{a+b}{2} = WP \quad (7)$$

WP: 각 기상상태 별 교통사고 발생 확률

마지막으로 다음에 발생할 교통사고 발생 건수를 예측 한다. 식(8)은 각 기상 상태 별 교통사고 예측 수를 구하는 식이다.

$$\text{각기상상태별교통사고예측수} = \sum_{i=1}^n WP(S_k) WM(S_i) \quad (8)$$

WM: 각 기상 상태별 교통사고 발생 건수 중 최대값

4. 기상 상태에 따른 교통사고 빈도수 예측 모델링 구현

4.1 기상 상태에 따른 교통사고 수 예측 확률

본 논문에서는 다음의 가정으로 기상상태에 따른 교통사고 발생 수를 예측하였다.

1. 기상 상태 중 맑은 날과 흐린 날 발생 할 교통사고 발생 수를 예측 하였다.
2. 1년 단위로 예측 하였다. 즉 2003년~2011년의 자료를 이용하여 2012년에 발생할 교통사고 수를 예측 하였다.
3. 예측된 자료와 실제 발생한 교통사고 수를 비교 분석하였다.

교통사고 데이터는 도로교통공단에서 매년 공개하는 통계자료를 이용하였다[8]. 표2는 교통사고 통계자료를 기상상태 별로 분석한 자료이다.

[Table 2] The number of traffic accidents

교통사고발생년도	교통사고 발생 건수		
	전체	맑음	흐림
2003	240,832	181,053	23,127
2004	220,775	182,021	14,133
2005	214,711	179,009	12,881
2006	213,745	177,447	12,487
2007	211,662	174,561	13,405
2008	215,882	186,143	10,384
2009	231,990	198,586	11,128
2010	226,878	185,850	14,035
2011	221,711	184,835	12,239
2012	223,656	186,840	10,850

4.1.1 교통사고 유형 (전체 교통사고)

전체 교통사고 발생 수를 분석하여 임계값의 범위와 상태를 정의 했다.

● 임계값 범위

$$S_1: 0 \sim 230,000 \quad S_2: 230,000 \sim 260,000$$

● 교통사고 발생 상태 (S_1)

$$S = \{S_1, S_2\}$$

초기 확률은 2010년, 2011년에 발생한 교통 사고 발생 건수를 식 (3)에 적용하여 얻는다. 식 (9)은 초기확률 식이다.

$$P(S_1:2, S_2:0) \quad P(1, 0) \quad (9)$$

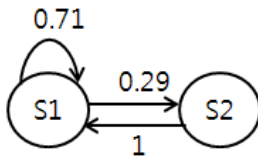
표1의 전체 교통사고 발생 수를 임계값 범위에 매핑하여 상태를 나열하면 다음과 같다.

$$S_1 \quad S_1 \quad S_2 \quad S_1 \quad S_1 \quad S_1 \quad S_1 \quad S_1 \quad S_2$$

각 상태가 다른 상태로 전이하는 전이 횟수를 구하고 식 (4)를 적용하여 전이확률로 계산한다. 계산된 데이터를 전이 행렬 표현 하면 식(10)이다.

$$\begin{matrix} S_1 & S_2 \\ S_1 & \begin{pmatrix} 0.71 & 0.29 \\ 1 & 0 \end{pmatrix} \\ S_2 & \end{matrix} \quad (10)$$

식 (10)의 상태 전이 다이어그램은 그림2 이다.



[Fig. 2] State transition diagram
(all traffic accidents)

교통사고 예측 확률은 초기확률과 전이행렬을 식(6)에 적용하여 구한다. 식(6)로 교통사고 발생 확률을 구할 수 있다.

$$(1 \ 0) \begin{pmatrix} 0.71 & 0.29 \\ 1 & 0 \end{pmatrix} = (0.71 \ 0.29) \quad (11)$$

식(11)에 의해 다음달에 발생할 교통사고 확률은 상태가 S_1 일 때 0.71로 가장 높다. 가장 높은 확률인 0.71을 교통사고 예측확률로 적용하였다.

4.1.2 교통사고 유형 (기상상태: 맑음)

기상 상태가 맑은 날의 교통사고 발생 수를 분석하여 임계값의 범위와 상태를 정의 했다.

● 임계값 범위

S_1 : 0~180,000 S_2 : 180,001~200,000

● 교통사고 발생 상태 (S_1)

$S = \{S_1, S_2\}$

초기 확률은 2010년, 2011년에 발생한 교통 사고 발생 건수를 식 (3)에 적용하여 얻는다. 식 (9)은 초기확률 식이다.

$$P(S_1:0, S_2:2) \quad P(0, 1) \quad (9)$$

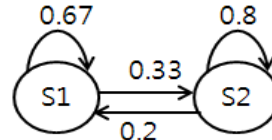
표1의 전체 교통사고 발생 수를 임계값 범위에 매핑하여 상태를 나열하면 다음과 같다.

$S_2 \ S_2 \ S_2 \ S_2 \ S_1 \ S_1 \ S_1 \ S_2 \ S_2$

각 상태가 다른 상태로 전이하는 전이 횟수를 구하고 식 (4)를 적용하여 전이확률로 계산한다. 계산된 데이터를 전이 행렬 표현 하면 식(12)이다.

$$\begin{matrix} S_1 & S_1 \\ S_1 & \begin{pmatrix} 0.67 & 0.33 \\ 0.2 & 0.8 \end{pmatrix} \\ S_2 & \end{matrix} \quad (12)$$

식 (12)의 상태 전이 다이어그램은 그림 3 이다.



[Fig. 3] State transition diagram
(Traffic accident occurred a sunny day)

교통사고 예측 확률은 초기확률과 전이행렬을 식(6)에 적용하여 구한다. 식(6)로 교통사고 발생 확률을 구할 수 있다.

$$(0 \ 1) \begin{pmatrix} 0.67 & 0.33 \\ 0.2 & 0.8 \end{pmatrix} = (0.2 \ 0.8) \quad (13)$$

식(13)에 의해 다음 달에 발생할 교통사고 확률은 상태가 S_2 일 때 0.8로 가장 높다. 가장 높은 확률인 0.8을 교통사고 예측확률로 적용하였다.

4.1.2 교통사고 유형 (기상상태: 흐림)

기상 상태가 흐린 날의 교통사고 발생 수를 분석하여 임계값의 범위와 상태를 정의 했다.

● 임계값 범위

S_1 : 0~14,000 S_2 : 14,001~25,000

● 교통사고 발생 상태 (S_1)

$S = \{S_1, S_2\}$

초기 확률은 2010년, 2011년에 발생한 교통 사고 발생 건수를 식 (3)에 적용하여 얻는다. 식 (14)은 초기확률 식이다.

$$P(S_1:1, S_2:1) \quad P(0.5, 0.5) \quad (14)$$

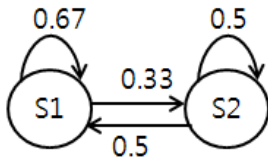
표1의 전체 교통사고 발생 수를 임계값 범위에 매핑하여 상태를 나열하면 다음과 같다.

$$S_1 \ S_2 \ S_1 \ S_1 \ S_1 \ S_1 \ S_1 \ S_2 \ S_2$$

각 상태가 다른 상태로 전이하는 전이 횟수를 구하고 식 (4)를 적용하여 전이확률로 계산한다. 계산된 데이터를 전이 행렬 표현 하면 식(15)이다.

$$\begin{matrix} & S_1 & S_2 \\ \begin{matrix} S_1 \\ S_2 \end{matrix} & \begin{pmatrix} 0.67 & 0.33 \\ 0.5 & 0.5 \end{pmatrix} \end{matrix} \quad (15)$$

식 (15)의 상태 전이 다이어그램은 그림4 이다.



[Fig. 4] State transition diagram
(Traffic accident occurred a cloudy day)

교통사고 예측 확률은 초기확률과 전이행렬을 식(6)에 적용하여 구한다. 식(6)로 교통사고 발생 확률을 구할 수 있다.

$$(0.5 \ 0.5) \begin{pmatrix} 0.67 & 0.33 \\ 0.2 & 0.8 \end{pmatrix} = (0.59 \ 0.41) \quad (16)$$

식(16)에 의해 다음 달에 발생할 교통사고 확률은 상태가 S_1 일 때 0.59로 가장 높다. 가장 높은 확률인 0.59을 교통사고 예측확률로 적용하였다.

4.2 기상 상태에 따른 교통사고 수 예측

표 3 은 마코프 프로세스가 예측한 2012년 발생할 기상 상태에 따른 교통사고 발생 예측확률이다.

[Table 3] According to the weather conditions the probability of traffic accidents

상태	전체 상태	기상상태 (맑음)	기상상태 (흐림)
예측확률	0.71	0.8	0.59

식(7)에 표3에 있는 각 기상 상태 별 예측 확률(WP)을 계산하면 다음과 같다.

기상상태 (맑음) : 0.76

기상상태 (흐림) : 0.65

표4는 식(8)을 적용하여 기상 상태 교통사고 발생 예측 건수이다.

[Table 4] Each weather conditions prediction by the number of traffic accidents

	예측확률 (WP)	각 기상 상태별 교통사고최고 발생건수	각 기상 상태별 교통사고 발생 예측 빈도수
기상상태 (맑음)	0.76	198,586	150,925
기상상태 (흐림)	0.65	23,127	15,033

본 논문에서 예측한 교통사고 발생 예측 빈도수와 실제로 2012년에 발생한 교통사고 발생 빈도수를 비교하면 다음의 표 5와 같다.

[Table 5] Comparison of the incidence and the actual occurrence prediction number of traffic accidents

	교통사고 발생 예측 빈도수	실제 발생한 교통사고 발생 빈도수 (2012년)	비율
기상상태 (맑음)	150,925	186,840	0.81
기상상태 (흐림)	23,127	10,850	0.72

5. 결론

기상 상태는 교통사고와 교통사고 사상자수에 큰 영향을 준다. 교통사고 사상자 수를 줄이기 위해서는 기상 상태에 따른 교통사고를 예측하는 것이 필요하다. 본 논문은 기상 상태에 따른 교통사고를 예측하는 모델링을 제안 하였다. 그리고 데이터 적용은 교통사고가 많이 발생하는 것을 기준으로 맑은 날과 흐린 날을 대상으로 하였다. 실제 데이터를 적용하여 본 논문에서 제안하는 모델링을 검증하였다. 검증 결과 실제 교통사고 발생 건수

와 70% 이상 유사한 결과를 얻었다.

본 모델링을 적용하면 매년 발생할 교통사고 예방 정책 수립에 도움이 될 것으로 예상된다.

향후에는 교통사고에 영향을 주는 악천후, 안개 등 다양한 기상 상태와 도로조건, 고속도로 등에 대한 연구를 진행할 예정이다.

REFERENCES

- [1] Analysis of Traffic Accident 2013 (Analysis of the characteristics of pedestrian traffic accident)(2013) <http://www.koroad.or.kr/>
- [2] Jong-Hak Lee, Hey-Jin Cho, "An Study on the Characters of Traffic Accidents in the Foul Weather According to Weather Condition", KOREAN SOCIETY OF CIVIL ENGINEERS, pp. 3867-3870, 2004.
- [3] Jean Andrey, Brian Mills, Jessica Vandermolen "A Temporal Analysis of Weather-related Collision Risk for Ottawa, Canada: 1990-1998", TRB, Paper 03-3488, 2002.
- [4] Chengyu Dai, "Exploration of Weather Impacts on Freeway Traffic Operations and Safety Using High-Resolution Weather Data", Master of Science in Civil and Environmental Engineering Portland State University, 2011.
- [5] Charles M. Grinstead, Introduction to Probability: Second Revised Edition, American Mathematical Society, pp. 405-416, 1997.
- [6] Young-Gab Kim, Young-kyo Baek, Hoh Peter In, Doo-Kwon Baik, A Probabilistic Model of Damage Propagation based on the Markov Process, Journal of KIISE, Vol. 33, No. 8, pp. 524-535, 2006.
- [7] OLIVER C. IBE, "Fundamentals of Applied Probability and Random Processes", Hantee media, 2008.
- [8] 2012 traffic accident statistics (2013 Edition) <http://www.koroad.or.kr/>

저자소개

정 영 석(Young-suk Chung)

[정회원]



- 2009년 : 공주대학교 멀티미디어 공학과(공학석사)
- 2013년 : 공주대학교 컴퓨터공학과 (공학박사)
- 2013 ~ 현재 : 대전보건 대학교 바이오정보과 겸임교수

<관심분야> : 시뮬레이션, 클라우드 컴퓨팅, 모바일 컴퓨팅, 정보보안

박 구 락(Rack-Koo Park)

[정회원]



- 1986년 : 중앙대학교 전기공학과 (공학사)
- 2002년 : 숭실대학교 전자계산학과 공학석사.(공학석사)
- 2000년 : 경기대학교 전자계산학과(이학박사)

· 1992 ~ 현재 : 공주대학교 컴퓨터공학부 교수

<관심분야> : 정보경영, 정보통신, 전자상거래

김 진 묵(Jin-Mook Kim)

[정회원]



- 2000년 : 배재대학교 컴퓨터공학과(공학석사)
- 2006년 : 광운대학교 컴퓨터과학과 (공학박사)
- 2006년 ~ 현재 : 선문대학교 IT 교육학부 부교수

<관심분야> : 정보보안, 유.무선 네트워크 보안, 클라우드 컴퓨팅 보안