

마코프 프로세스를 적용한 범죄 발생 예측 방법에 관한 연구

정영석*, 정진영**

A Study of the Prediction of Incidence of Crime using Markov process

Young-Suk Chung*, Jin-Young Jung**

요약

현대 사회는 다양한 범죄들이 발생하고 있고, 범죄를 예방하기 위한 연구가 진행되고 있다. 기존의 범죄에 관련된 연구들은 범죄가 발생하는 공간과 지리정보를 분석하거나, 범죄자들의 범죄 유형을 분석하는 연구들이 진행되어 왔다. 그러나 기존의 연구들은 지리적, 심리학적인 연구를 통해 범죄가 발생하는 지역과 동기들을 분석하여 범죄를 예방하기 위한 연구들이 대부분이다. 본 논문에서는 마코프 프로세스를 도입하여 범죄를 예측하기 위한 모델링을 제시한다. 여러 범죄 중 살인, 공무원 범죄, 폭력의 범죄 발생 건수를 사용하여 시간에 따른 범죄 발생 건수를 예측하였다. 본 논문에서 제시한 범죄 예측 모델링에서 사용될 범죄 발생 평균값에 범죄가 발생한 기간에 발생한 범죄 발생 건수의 전체 평균값, 1년 평균값, 최근 평균값으로 분류하여 어느 것이 예측 정확도를 높일 수 있는 지 비교하였고, 최근 평균값을 적용하는 것이 범죄 발생 예측정확도를 높일 수 있음을 확인하였다.

▶ Keyword : 마코프 프로세스, 시뮬레이션, 미래 예측 모델링

Abstract

Modern society is experiencing a variety of crimes, and to prevent crime is being studied. Existing studies related to the crime of crimes that occur on spatial analysis and geographic information, or to analyze the type of criminal offense of studies have been conducted, However the existing studies of the geographical and psychological crime that occurs throughout the study area and by analyzing the motives for the crime prevention research is the most.

In this paper, we introduce Markov processor model for predicting the crime is present. Of several crimes, murder, government official crimes, the incidence of violent crime has occurred over

• 제1저자 : 정영석 • 교신저자 : 정진영

• 투고일 : 2011. 11. 14, 심사일 : 2011. 11. 28, 게재확정일 : 2011. 12. 28.

* 공주대학교 컴퓨터공학과(Dept. of Computer Science & Engineering, Kongju national University)

** 대전보건대학교 바이오정보과(Dept. of Bio Information)

time by using the predicted incidence of crime. Presented in this paper, predictive modeling is used in a crime occurred in the average duration of the overall average number of crimes that occurred in the one-year average, which recently labeled as the average prediction was compared to if you can increase the likelihood, recent average to apply to increase the probability of the prediction that crime have been investigated.

▶ Keyword : Markov Process, Simulation, Predictive modeling,

I. 서 론

오늘날 산업과 과학기술의 발전으로 인류의 삶의 질은 점점 더 향상되고 있다. 그러나 실업, 빈곤, 범죄 등 다양한 사회 문제들이 발생 하고 있다. 그 중 범죄는 인간이 활동하고 거주하는 모든 지역에서 발생하고 있고, 범죄로 인한 피해도 증가 추세에 있다. 그러므로 범죄를 예방하기 위해 다양한 연구가 이루어지고 있다. 범죄자와 피해자의 거주지 사이에 공간적 분포의 패턴을 연구하여, 특정한 패턴의 원인을 찾는 것이 연구 되었고[1][2], GIS(Geographic Information System)지도와 데이터베이스를 이용하여 범죄의 공간적 분포를 분석하여 시각화 하는 연구가 이루어졌다[3]. 그리고 범죄자들의 범죄 경력을 바탕으로 유사한 범죄를 지속적으로 저지르는지, 아니면 다양한 범죄들을 저지르는 지에 대한 범죄 유형전환에 관한 연구도 진행되었다[4]. 이와 같이 범죄에 대한 지리적, 심리적 요인 분석에 관련된 연구는 활발하게 진행되고 있으나, 범죄를 예방하기 위한 범죄 예측에 관련된 연구는 부족한 실정이다. 그러므로 본 논문은 다양한 분야의 예측에 사용되는 마코프 프로세스 이론을 이용하여 범죄 발생 건수를 예측하기 위한 연구를 진행하고자 한다. 이를 위해 마코프 프로세스를 이용하여 범죄 발생 예측을 위한 모델링을 구성해 보았다. 그리고 실제 범죄 발생 건수를 적용해 예측값을 구하였다. 그리고 그 예측값과 실제값을 비교해 보았다. 추가로 범죄 발생 건수의 예측을 높이는 방법에 대해서도 기술하였다. 본 논문의 구성은 다음과 같다. 2장에서 관련연구인 마코프 프로세스에 대해 논의하고 3장에서는 마코프 프로세스를 이용한 범죄 예측 모델링에 대해 제시한다. 4장에서는 마코프 프로세스 모델링을 바탕으로 실제 범죄 데이터를 이용하여 예측하고 그 값을 실제 발생 건수와 비교해 본다. 5장에서는 결론 및 향후 연구 과제에 대해 논의 한다.

II. 관련 연구

1. 관련연구

1.1 마코프 프로세스

마코프 프로세스는 마코프 성질을 가진 이산 확률과정으로서 러시아의 수학자인 안드레이 마코프의 이름에서 왔다. 마코프 프로세스는 어떤 사건에서 직전의 상태가 현재의 상태로 영향을 주고, 과거의 상태에는 전혀 영향을 받지 않는 확률 과정을 가정한 것이다. 마코프 프로세스는 과거의 동적 특성을 분석하여 미래에 있을 변화를 예측하기 위한 수학적 기법이다[5].

마코프 프로세스는 상태간의 전이가 오로지 이전 n 개의 상태에 의존하는 프로세스를 의미한다. 이때 이 모델을 n 차원 모델이라고 하는데 n 은 다음 상태를 결정하는데 영향을 주는 상태들의 수를 의미한다[6]. 마코프 프로세스를 $X(t)$ 라 하면 임의의 시간 $t_1 < t_2 < \dots < t_k < t_{k+1}$ 에 대해 $X(t)$ 가 이산 값이면

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

식(1)로 기술할 수 있고, $X(t)$ 가 연속값이면

$$P[a < X(t_{k+1}) = b | X(t_k) = x_k, \dots, X(t_1) = x_1] \quad (2) \\ = P[a < X(t_{k+1}) = b | X(t_k) = x_k]$$

로 기술된다. 위의 식(1), 식(2)에서 t_k 는 현재, t_{k+1} 은 미래, t_1, \dots, t_{k-1} 은 과거의 시점이다. 마코프 프로세스의 값이 이산값이면 마코프 체인이라고 정의 한다. 마코프 체인은 t 가 이산 적이나 연속적이냐 에 따라 이산시간 마코프 프로세스와 연속시간 마코프 프로세스로 나눌 수 있다. 마코프 프로세스는 가능한 상태들의 집합인 상태집합, 모델링의 초기화 확률 벡터인 초기 확률, 각 상태간의 전이확률로 구성된 전이 행렬로 구성되어 있다[6,7]. 미래를 예측할 수 있는 마코프 프로세스의 특성은 다양한 분야에서 활용되고 있다. 마코프 프로세

스를 이용하여 정류장 사이를 전이확률을 이용한 행렬 표를 생성하여 버스지체 시간을 예측하기 위한 연구가 있었고[8], 고무 차륜 AGT 차량시스템의 운행상태를 모형화 하여 신뢰도와 가용성을 예측하는 연구가 있었고[9], 공동주택내의 쾌적한 실내 환경을 유지하기 위해 재실자의 행동을 확률적으로 예측하는 연구에 사용되는데[10]등 다양한 분야에 활용되고 있다.

III. 본 론

1. 범죄예측 모델링

마코프 프로세스를 이용한 범죄 예측 모델링을 구성해 보면 그림 1과 같다.

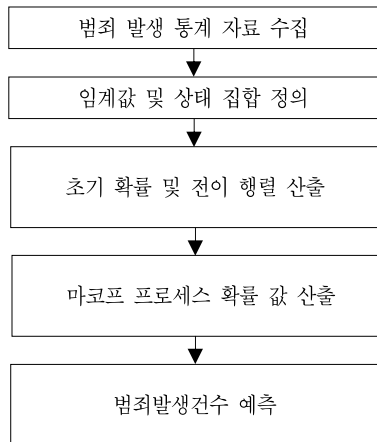


그림 1. 마코프 프로세스를 이용한 범죄 예측 모델링
Fig. 1. Crime prediction modeling using Markov processes

대검찰청이 매년 발표하는 범죄 발생 통계 자료를 1년 단위로 수집한 후 분석하여 임계값을 설정하고, 임계값을 기준으로 각 상태들을 정의한다. 정의된 상태들을 범죄 발생 통계자료와 매핑시켜 상태 집합을 생성한다. 생성된 상태집합을 바탕으로 초기 확률과 전이행렬을 생성하고, 이것을 마코프 프로세스를 이용한 범죄 예측 모델링에 입력하여 범죄 발생 예측 확률을 구한 후, 예측 확률과 범죄 발생 평균값을 이용하여 미래에 발생할 범죄 발생 건수를 예측 한다.

1.1 범죄 예측 모델링 구현

대검찰청에서는 매년 전년도에 범죄 통계자료를 공개한다. 공개된 범죄 통계자료 중 살인, 공무원 범죄, 폭력을 대상으

로 범죄 발생 건수를 예측 하였다. 범죄 발생 건수를 분석하여 상태에 따른 임계값을 설정하여 상태집합을 정의하고, 초기 확률, 전이 행렬을 계산한 후, 초기 확률과 전이 행렬을 이용하여 마코프 프로세스 확률 값을 구한다[6]. 산출된 확률 값을 범죄 발생 예측 식에 적용하여 범죄 발생 예측값을 계산한다.

◆ 상태 집합(S): 범죄 예측 모델링에서 상태란 한 범죄가 발생하는 빈도수의 범위 말하며, 범죄 통계 자료를 바탕으로 적절한 임계값을 설정하여 상태들을 집합으로 정의하였다.

◆ 초기 확률: 한 범죄가 초기상태에서 가질 수 있는 범죄 발생 확률로서, 최근에 발생한 범죄의 발생 상태를 이용하여 식(3)과 같이 정의한다.

$$P(S_1, S_2, \dots, S_n) = P\left(\frac{a}{F}, \frac{b}{F}, \dots, \frac{c}{F}\right) \quad (3)$$

여기서 a, b, c 는 각 상태(S_1, S_2, \dots, S_n)의 범죄 발생 횟수이고 F 는 a, b, c 의 합이다. 또한 초기 확률의 총 합은 1이 되어야하는 조건(4)을 만족한다.

$$\sum_{i=1}^N P(S_i) = 1 \quad (S\text{는 상태}) \quad (4)$$

◆ 전이행렬: 정의된 상태간의 전이 상태를 확률로 나타낸 것이다. 각 범죄 발생 데이터들을 상태 집합과 매핑 하여 상태들을 열거한다. 그리고 열거된 하나의 상태에서 다른 상태로의 전이 회수를 구한 후, 이를 전이 행렬로 나타낸 것이다. 각 행의 합은 1인 조건을 만족한다. 식(5)의 P는 전이행렬로서 조건(6)을 만족한다.

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} & \dots & P_{1n} \\ P_{21} & P_{22} & P_{23} & \dots & P_{2n} \\ \dots & \dots & \dots & P_{ij} & \dots \\ P_{n1} & P_{n2} & P_{n3} & \dots & P_{nn} \end{pmatrix} \quad (5)$$

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

$$\sum_{j=1}^n P_{ij} = 1, i = 1, 2, \dots, n \quad (6)$$

◆ 범죄 발생 확률: 마코프 프로세스로 식(7)로 정의된다.

$$P(S_k) = \sum_{i=1}^n P(S_i)P_{ik} \quad (7)$$

$P(S_i)$: 초기 확률 P_{ik} :전이행렬

◆ 범죄 발생 예측 식 : 기존 논문은 각 위험상태의 중간값을 사용하여 마코프 프로세스의 예측값을 구하였으나[6],본 논문에서는 범죄 건수를 예측할 때 다양한 평균값을 적용하였다.

$$\text{범죄 발생 예측 건수} = \sum_{i=1}^n P(S_i) WM(S_i) \quad (8)$$

$$\text{범죄 발생 예측 건수} = \sum_{i=1}^n P(S_i) YM(S_i) \quad (9)$$

$$\text{범죄 발생 예측 건수} = \sum_{i=1}^n P(S_i) RM(S_i) \quad (10)$$

n : 범죄발생 상태 집합의 상태 수

$P(S_i)$: 범죄 발생 확률,

$WM(S_i)$: 범죄 발생 건수의 전체 평균 값

(본 논문에서는 3년간 발생한 범죄 발생 건수 평균값)

$YM(S_i)$: 최근 1년 간 발생한 범죄 발생 건수의 평균값

$RM(S_i)$: 최근 6개월간 발생한 범죄 발생건수의 평균값

1.1.1 범죄 유형(살인)

표 1은 2006~2008년 발생한 살인 범죄 발생 건수를 나타낸 것으로, 범죄 발생 건수를 분석하여 범죄 발생 상태를 정의하였고, 상태에 맞는 임계값의 범위를 설정하였다.

표 1. 범죄 발생 건수(murder)
Table 1.The incidence of crime(murder)

	2006년	2007년	2008년
1월	74	96	66
2월	69	83	50
3월	82	104	69
4월	81	89	105
5월	105	96	116
6월	96	91	92
7월	110	103	100
8월	114	98	116
9월	85	104	116
10월	92	108	105
11월	78	82	91
12월	78	70	94

◆ 임계값 범위

S_1 :0~70 S_2 :71~110 S_3 :111~150

◆ 범죄 발생 상태 (S)

$S = \{S_1, S_2, S_3\}$

식 (3)을 이용하여 2008년 8월~12월 까지 발생한 범죄

발생 건수 표 1을 이용하여 초기 확률 식(11)을 구한다.

◆ 범죄발생건수 : 116, 116, 105, 91, 94
 $S_3 \ S_3 \ S_2 \ S_2 \ S_2$

◆ 초기 확률 : $P(S_1:0, S_2:3, S_3:2)$
 $P(0, 0.6, 0.4)$ (11)

표 1의 범죄 발생 건수를 임계값의 범위에 매핑하여 상태들을 나타낸다.

$S_2 \ S_1 \ S_2 \ S_2 \ S_2 \ S_2 \ S_2 \ S_3 \ S_2 \ S_2 \ S_2 \ S_2$
 $S_2 \ S_2 \ S_2 \ S_2 \ S_2 \ S_2 \ S_2 \ S_2 \ S_2 \ S_2 \ S_2 \ S_1$
 $S_1 \ S_1 \ S_1 \ S_2 \ S_3 \ S_2 \ S_2 \ S_3 \ S_3 \ S_2 \ S_2 \ S_2$

각 상태 (S_1, S_2, S_3) 가 다른 상태로 전이하는 전이확률을 식(5)을 이용하여 전이확률로 만든 후, 전이행렬로 표현하면 식(12)로 표현된다.

$$\begin{matrix} & S_1 & S_2 & S_3 \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \end{matrix} & \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.08 & 0.81 & 0.11 \\ 0 & 0.75 & 0.25 \end{pmatrix} \end{matrix} \quad (12)$$

그리고 식(12)을 상태전이 다이어그램으로 표현하면 그림 2와 같다.

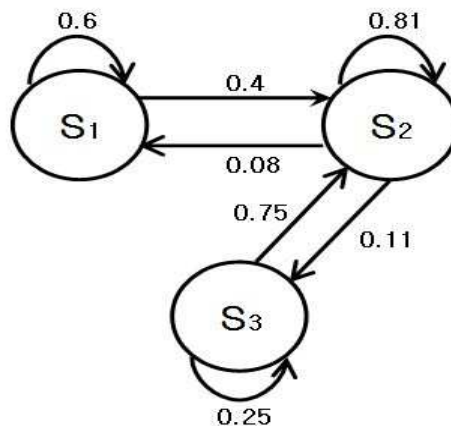


그림 2 범죄에 대한 상태 다이어그램(살인)
Fig. 2. The state diagram for a crime (murder)

식(3), 식(5)을 이용하여 초기 확률과 전이행렬을 구한 후 식(7)을 적용하면 범죄 발생 확률을 구할 수 있다.

$$(0 \ 0.6 \ 0.4) \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.08 & 0.81 & 0.11 \\ 0 & 0.75 & 0.25 \end{pmatrix} \quad (13)$$

$$= (0.05 \ 0.79 \ 0.16)$$

구해진 식(13)에 의해 다음 달 범죄 발생 확률은 상태가 S_2 일 때 0.75로 가장 높다. 다음 달에 발생할 범죄 발생 건수를 예측하기 위해 발생 확률 값에 살인 범죄 발생건수의 전체 평균값, 최근1년 간 평균값, 최근 평균값을 적용하였다.

◆ 전체 평균값 :

식(8)에 2006년~2008년에 발생한 살인 범죄 발생건수의 평균값을 적용한다.

$$WM(S_i) = 91.88\text{건}$$

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) WM(S_i) \\ &= 0.75 \times 91.88 \\ &= 72.59 \end{aligned}$$

◆ 최근 1년간 평균값

식(9)에 2008년에 발생한 살인 범죄 발생건수의 평균값을 적용한다.

$$YM(S_i) = 93.33$$

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) YM(S_i) \\ &= 0.79 \times 93.33 \\ &= 73.733 \end{aligned}$$

◆ 최근 평균값

식(10)에 2008년 7월~ 2008년12월(6개월)에 발생한 살인사건의 평균값을 적용한다.

$$RM(S_i) = 103.66$$

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) RM(S_i) \\ &= 0.79 \times 103.66 \\ &= 81.89 \end{aligned}$$

살인 범죄 발생 건수를 예측한 결과 전체 평균값일 때 약 73건, 최근 1년 평균값일 때 약 74건, 최근 평균값일 때 약 82건이 예측되었고 실제 2009년 1월 달에 발생한 살인 사건 발생건수는 79 건 이었다.

1.1.2 범죄 유형(공무원 범죄)

표 2는 2006~2008년 발생한 공무원 범죄 발생 건수를 나타낸 것으로, 범죄 발생 건수를 분석하여 범죄 발생 상태를 정의 하였고, 상태에 맞는 임계값의 범위를 설정하였다.

표 2 범죄 발생 건수(공무원 범죄)

Table 2 The incidence of crime(government official crime)

	2006년	2007년	2008년
1월	119	117	122
2월	130	121	83
3월	116	99	96
4월	90	96	95
5월	119	117	102
6월	122	154	110
7월	123	131	174
8월	124	139	97
9월	151	150	135
10월	118	135	194
11월	166	169	126
12월	155	159	158

◆ 임계값 범위

$$S_1:0\sim100 \ S_2:101\sim150 \ S_3:151\sim200$$

◆ 범죄 발생 상태 (S)

$$S = \{S_1, S_2, S_3\}$$

식 (3)을 이용하여 2008년 8월~12월 까지 발생한 범죄 발생 건수 표 1을 이용하여 초기 확률 식(14)을 구한다.

◆ 범죄발생건수 : 97, 135, 194, 126, 158

$$S_1, S_2, S_3, S_2, S_3$$

◆ 초기 확률 : $P(S_1:1, S_2:2, S_3:2)$

$$P(0.2, 0.4, 0.4) \quad (14)$$

각 상태 (S_1, S_2, S_3) 가 다른 상태로 전이하는 전이확률을 식(5)을 이용하여 전이확률로 만든 후, 전이행렬로 표현하면 식(15)로 표현된다.

$$\begin{matrix} S_1 & S_2 & S_3 \\ S_1 & \begin{pmatrix} 0.43 & 0.57 & 0 \\ 0.15 & 0.5 & 0.35 \\ 0.13 & 0.62 & 0.25 \end{pmatrix} \\ S_2 & \\ S_3 & \end{matrix} \quad (15)$$

식(3), 식(5)을 이용하여 초기 확률과 전이행렬을 구한 후 식(7)을 적용하면 범죄 발생 확률을 구할 수 있다.

$$\begin{matrix} (0.2 & 0.4 & 0.4) & \begin{pmatrix} 0.43 & 0.57 & 0 \\ 0.15 & 0.5 & 0.35 \\ 0.13 & 0.62 & 0.25 \end{pmatrix} \\ & & & \end{matrix} \quad (16)$$

$$= (0.2 \quad 0.56 \quad 0.24)$$

구해진 식(16)에 의해 다음 달 범죄 발생 확률은 상태가 S_2 일 때 0.56로 가장 높다. 다음 달에 발생할 범죄 발생 건수를 예측하기 위해 발생 확률 값에 살인 범죄 발생건수의 전체 평균값, 최근1년 간 평균값, 최근 평균값을 적용하였다.

◆ 전체 평균값 :

식(8)에 2006년~2008년에 발생한 공무원 범죄 발생건수의 평균값을 적용한다.

$$WM(S_i) = 128.1 \text{ 건}$$

$$\begin{aligned}
 \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) WM(S_i) \\
 &= 0.56 \times 128.1 \\
 &= 71.74
 \end{aligned}$$

◆ 최근 1년간 평균값

식(9)에 2008년에 발생한 공무원 범죄 발생건수의 평균값을 적용한다.

$$YM(S_i) = 124.3$$

$$\begin{aligned}
 \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) YM(S_i) \\
 &= 0.56 \times 124.3 \\
 &= 69.63
 \end{aligned}$$

◆ 최근 평균값:

식(10)에 2008년 7월~ 2008년12월(6개월)에 발생한 공무원 범죄 발생 건수의 평균값을 적용한다.

$$RM(S_i) = 147.3$$

$$\begin{aligned}
 \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) RM(S_i) \\
 &= 0.56 \times 147.3 \\
 &= 82.51
 \end{aligned}$$

공무원 범죄 발생 건수를 예측한 결과 전체 평균값일 때 약 72건, 최근 1년 평균값일 때 약 70건, 최근 평균값일 때 약 83건이 예측되었고 실제 2009년 1월 달에 발생한 살인 사건 발생건수는 101 건 이었다.

1.1.3 범죄 유형(폭력)

표 3은 2006~2008년 발생한 폭력 범죄 발생 건수를 나타낸 것으로, 범죄 발생 건수를 분석하여 범죄 발생 상태를 정의 하였고, 상태에 맞는 임계값의 범위를 설정하였다.

표 3. 범죄 발생 건수(폭력)
Table 3.The incidence of crime (violence crime)

	2006년	2007년	2008년
1월	21,394	20,482	19,577
2월	19,262	16,711	15,321
3월	20,558	19,565	18,638
4월	19,208	20,671	20,815
5월	21,800	22,306	21,800
6월	21,098	22,199	23,718
7월	21,231	24,213	24,544
8월	20,558	23,312	21,979
9월	22,155	19,582	24,996
10월	22,621	24,609	24,506
11월	22,155	22,047	22,046
12월	19,150	19,561	23,559

◆ 임계값 범위

S_1 :0~18,000 S_2 :18,001~21,000 S_3 :21,001~24,000
 S_4 :24,001~27,000

◆ 범죄 발생 상태 (S)

$$S = \{S_1, S_2, S_3, S_4\}$$

식 (3)을 이용하여 2008년 8월~12월까지 발생한 범죄 발생 건수 표 1을 이용하여 초기 확률 식(17)을 구한다.

◆ 범죄발생건수 : 21,979, 24,996, 24,506, 22,046, 23,559

$$S_3 \quad S_4 \quad S_4 \quad S_3 \quad S_3$$

◆ 초기 확률 : $P(S_1: 0, S_2: 0, S_3: 3, S_4: 2)$

$$P(0, 0, 0.6, 0.4) \quad (17)$$

각 상태 $\{S_1, S_2, S_3, S_4\}$ 가 다른 상태로 전이하는 전이 확률을 식(5)을 이용하여 전이확률로 만든 후, 전이행렬로 표현하면 식(18)로 표현된다.

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.15 & 0.46 & 0.31 & 0.08 \\ 0 & 0.33 & 0.47 & 0.2 \\ 0 & 0 & 0.8 & 0.2 \end{pmatrix} \quad (18)$$

식(3), 식(5)을 이용하여 초기 확률과 전이행렬을 구한 후 식(7)을 적용하면 범죄 발생 확률을 구할 수 있다.

$$(0 \ 0 \ 0.6 \ 0.4) \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.15 & 0.46 & 0.31 & 0.08 \\ 0 & 0.33 & 0.47 & 0.2 \\ 0 & 0 & 0.8 & 0.2 \end{pmatrix} \quad (19)$$

$$= (0 \ 0.2 \ 0.6 \ 0.2)$$

구해진 식(19)에 의해 다음 달 범죄 발생 확률은 상태가 S_2 일 때 0.6으로 가장 높다. 다음 달에 발생할 범죄 발생 건수를 예측하기 위해 발생 확률 값에 살인 범죄 발생건수의 전체 평균값, 최근1년 간 평균값, 최근 평균값을 적용하였다.

◆ 전체 평균값 :

식(8)에 2006년~2008년에 발생한 폭력 범죄 발생건수의 평균값을 적용한다.

$$WM(S_i) = 21,332 \text{건}$$

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) WM(S_i) \\ &= 0.6 \times 21,332 \\ &= 12,799 \end{aligned}$$

◆ 최근 1년간 평균값

식(10)에 2008년에 발생한 폭력 범죄 발생건수의 평균값을 적용한다.

$$YM(S_i) = 21,792$$

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) YM(S_i) \\ &= 0.6 \times 21,792 \\ &= 13,075 \end{aligned}$$

◆ 최근 평균값

식(10)에 2008년 7월~ 2008년12월(6개월)에 발생한 폭력 범죄 발생 건수의 평균값을 적용한다.

$$RM(S_i) = 23,605$$

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) RM(S_i) \\ &= 0.6 \times 23,605 \\ &= 14,163 \end{aligned}$$

폭력 범죄 발생 건수를 예측한 결과 전체 평균값일 때 12,799건, 최근 1년 평균값일 때 13,075건, 최근 평균값일 때 14,163건이 예측되었고 실제 2009년 1월 달에 발생한 살인 사건 발생건수는 17,797건 이었다.

IV. 평균값의 변화에 따른 범죄 발생 건수 분석

본 연구는 살인, 공무원 범죄, 폭력의 범죄를 마코프 프로세스 예측 모델링에 적용하여 범죄 발생 확률을 산출 하였다. 그리고 각각의 범죄 발생 건수를 예측하기 위해서 범죄 예측 모델링의 평균값에 변화를 주었고 그 결과를 표 4로 나타내었다.

표 4. 범죄 발생 예측 건수와 실제 발생 건수 비교
Table 4. To compare predict the number of crime and the actual of crime

구분	평균분류	범죄발생 예측건수	실제발생건수 (2009년1월)	차 이
살인	전체평균	73	79	6
	최근 1년 평균	74	79	5
	최근평균	82	79	3
공무원 범죄	전체평균	72	101	29
	최근 1년 평균	70	101	31
	최근평균	83	101	18
폭력	전체평균	12,799	17,797	4,998
	최근 1년 평균	13,075	17,797	4,722
	최근평균	14,163	17,797	3,634

전체평균은 2006년~2008년 사이에 발생한 범죄발생건수의 평균값이고, 최근 1년 평균은 2008년 발생한 범죄발생건수의 평균값이고, 최근 평균은 2008년 7월~12월(6개월)까지 발생한 범죄 발생건수의 평균값이다. 범죄 발생건수의 평균값에 변화를 주자 범죄 발생 예측 건수가 변화는 것으로 나타났고, 그중에서도 최근 평균값인 6개월을 적용하였을 때 실제 발생건수와 근접함을 알 수 있다.

V. 결 론

본 논문은 미래 예측 모델인 마코프 프로세스를 이용한 범죄 예측 모델링을 제시하였고, 살인, 공무원 범죄, 폭력 범죄를 사용하여 예측값을 계산해 보았다. 마코프 프로세스를 이용하여 연구기간의 다음 달에 발생할 범죄 발생 확률을 구하였고 범죄 발생건수 예측에 적용할 평균값으로 연구 대상 기간에 발생한 범죄 발생 건수의 전체 평균값, 최근 1년간 발생한 범죄 발생 건수의 평균값, 최근 6개월간 평균값으로 나누어 미래에 발생할 범죄 발생 건수를 예측해 보았다. 그 결과 최근에 발생한 범죄 발생 건수 평균값을 적용하는 것이 실제 발생건수와 근접함을 알 수 있었다. 본 논문에서 사용된 범죄 예측 모델링을 적용하면 가까운 미래에 발생할 범죄의 발생 건수를 예측 할 수 있으므로, 범죄에 대한 예방정책 수립에 도움을 줄 것으로 예상된다. 그러나 본 논문에서 적용한 마코프 프로세스로 나온 예측값은 먼 미래를 예측 할 수 없고, 사 이크 패스, 연쇄 살인범, 문지 마 살인 등 특별한 범죄의 경우 예측할 수 없다. 그러므로 향후에는 먼 미래를 예측하고, 특별한 범죄들도 예측이 가능 하도록 연구를 진행할 예정이다.

참고문헌

[1] Brown, M.A., "Modelling the Spatial Distribution of Suburban Crime," *Economy Geography*, Vol. 58, No3, pp. 247-261, July 1982.

[2] Kamber, T., Mollenkopf, H., and Ross, A. "Crime, Space, and Place : An Analysis of Crime Patterns in Brooklyn", in Goldsmith V., Mguire G., Mollenkopf, H. and Ross, A(eds.), *Analyzing Crime Patterns: Frontiers of Practice* Sage, pp121-136, 2000.

[3] Hwang Sun-Young, Hwang Chul-sue, "The Spatial Pattern Analysis Of Urban Crimes Using GIS : The Case of Residential Burglary", *Journal of Korea Planners Association*, Vol 38, No1, pp 53-56, 2, 2003.

[4] Park Cheol-Hyun, "Specialization in Criminal Career : Markov-Chain Analysis" *Korean Criminological Review*, pp. 243-273, 3, 2003.

[5] Charles M. Grinstead, "Introduction to Probability: Second Revised Edition", *American Mathematical Society*, pp405-406, 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*, Vol33, No8, pp.524-535, 8, 2006.

[7] Jung-Min Park, Koo-Rack Park, "A Study on the Prediction of Crime Probability by month using Markov Chains", *Journal of The Korea Knowledge Information Technology Society*, 2011. 2, Vol 6, No 1, pp. 69-77.

[8] Seung-Hun Lee, Byeong-Sup Moon, Bum-Jin Park , "The Bus Delay Time Prediction Using Markov Chain", *The Journal of The Korea Institute of Intelligent Transport Systems*, pp.1-10, 6, 2009.

[9] Ha Chen-Soo, Han Seok-Youn, "Reliability Evaluation of AGT Vehicle System Using Markov Chains" 2003 Autumn Conference & Annua Meeting of the Korean Society for Railway, pp 91-96, 10, 2004.

[10] Kim Young-Jin, Park Cheol-Soo, "Prediction of Occupant's Presence in Residential Apartment Buildings using Markov Chain" *Korea Institute of Architectural Sustainable Environment and building System*. 2008 autumn conference, pp116~121, 2008.

저 자 소 개



정 영 석

1998 : 서남대학교 물리학과 이학사.
2000 : 배재대학교 대학원 물리학과 이
학석사.
2009 : 공주대학교 대학원 멀티미디어
공학과 공학석사.
현 재 : 공주대학교 대학원 컴퓨터공학과
박사과정.
관심분야 : 시뮬레이션, 클라우드 컴퓨팅,
패턴인식,
Email : merope@kongju.ac.kr



정 진 영

1992 : 한남대학교 전자계산학과 공학사.
1994 : 한남대학교 대학원 컴퓨터공학과
공학석사.
2002 : 한남대학교 대학원 컴퓨터공학과
공학박사.
현 재 : 대전보건대학교 바이오정보과
교수.
관심분야 : 멀티미디어 문서처리(XML),
객체지향 모델링 및 방법론,
분산시스템 및 실시간 시스
템, 시뮬레이션 등
Email : jyjung@hit.ac.kr