

시간 상태 변화를 적용한 범죄 발생 예측에 관한 연구

박 구 락*

A Study of the Probability of Prediction to Crime according to Time Status Change

Koo-Rack Park

요 약

현대 사회의 각 분야는 산업화와 과학기술의 발전으로 빠르게 변화한다. 그러나 빠른 사회 변화의 부작용으로 다양한 문제가 발생하고 있는데, 그 중 범죄는 큰 문제이다. 본 논문은 범죄를 예측하기 위한 모델로 마코프 체인을 적용한 범죄 예측 모델링을 제안한다. 기존의 마코프 체인 모델링은 한 사건의 전체 상태만으로 미래 예측 확률을 구하였으나, 본 논문은 사건 발생 확률 예측을 높이기 위해 전체 상태 예측 확률과 최근 상태 예측 확률로 나누었다. 그리고 전체 상태 예측 확률과 최근 상태 예측 확률의 평균값을 적용하여 미래 예측 확률 모델링으로 구현했다. 데이터는 범죄 발생 건수를 적용하였다. 그 결과 전체 상태만을 대상으로 예측확률을 적용 하였을 때 보다, 전체 상태와 최근상태로 나누어 확률 값을 구한 후, 그 평균값을 예측 확률로 적용하였을 때, 범죄 발생 예측에 근접하다는 결론을 얻었다.

▶ Keywords : 미래 예측, 마코프 체인, 시뮬레이션, 범죄 통계

Abstract

Each field of modern society, industrialization and the development of science and technology are rapidly changing. However, as a side effect of rapid social change has caused various problems. Crime of the side effects of rapid social change is a big problem. In this paper, a model for predicting crime and Markov chains applied to the crime, predictive modeling is proposed. Markov chain modeling of the existing one with the overall status of the case determined the probability of predicting the future, but this paper predict the events to increase the probability of occurrence probability of the prediction and the recent state of the entire state was divided by the probability of the prediction. And the whole state and the probability of the prediction and the recent state by

•제1저자 : 박구락

•투고일 : 2013. 3. 16, 심사일 : 2013. 4. 22, 게재확정일 : 2013. 5. 21.

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

applying the average of the prediction probability and the probability of the prediction model were implemented. Data was applied to the incidence of crime. As a result, the entire state applies only when the probability of the prediction than the entire state and the last state is calculated by dividing the probability value. And that means when applied to predict the probability, close to the crime was concluded that prediction.

▶ Keywords : Predictive modeling, Markov chains, Simulation, Crime statistics

I. 서론

유현대사회는 물질적 풍요와 과학기술의 발전으로, 사회의 각 분야에 대한 변화 속도가 과거와 다르게 빠르게 변화 하고 있다. 그에 따라 사회의 각 분야의 변화에 대한 예측을 하고 대비하는 것에 대한 필요성이 증가 하고 있다[1]. 사회 변화를 예측하기 위한 연구로 회귀분석, 시스템 다이내믹스, 마코프 체인 등이 있고, 본 논문은 미래 예측 방법 중 하나인 마코프 체인을 이용하였다. 마코프 체인은 시간에 따라 발생할 사건을 예측하는 기법으로 다양하게 사용되고 있다. 데이터는 현대 사회의 큰 문제 중의 하나인 범죄를 대상으로 하였다. 범죄데이터는 대검찰청에서 매년 발표하는 범죄 발생 건수를 월별로 수집하여, 시간에 따른 범죄의 발생을 예측하는데 적용하였다. 기존에도 마코프 체인을 사용하여 미래에 발생할 범죄를 예측하는 연구가 있었으나[2], 본 논문에서는 예측 확률을 높이기 위해, 마코프 체인 확률을 사건이 발생한 기간을 전체를 대상으로 하는 전체 확률과 최근 발생한 사건을 대상으로 하는 최근 확률로 나누어서 전체 확률과 최근확률의 평균값을 사건 발생 예측 확률로 사용하였다. 본 논문의 구성은 다음과 같다. 2장 관련연구에서 마코프 체인에 대해 알아보고 3장에서 본 논문에 적용될 마코프 체인을 적용한 범죄 예측 모델링을 제시하고 4장에서 실제 범죄 데이터를 적용하여 범죄를 예측하고 그 값을 실제 발생 데이터와 비교한다. 5장에서는 결론 및 향후 연구 과제에 대해 논의 한다.

II. 관련 연구

2.1 범죄 예측 연구

범죄는 범죄를 당한 피해자 뿐 만 아니라 피해자 주변인들

에게도 큰 영향을 미친다. 범죄를 분석하고 예방하려는 다양한 연구가 진행되었다. 범죄자들이 같은 범죄를 반복적으로 일으키는 것을 전문화라고 하는데, 전문화를 통해 범죄를 분석하는 방법에 대해 연구가 진행되었다[3]. 범죄가 발생하는 환경을 조사하고, GIS(Geographic Information System)의 공간 분석기능과 공간통계 분석 방법을 통합해서 범죄가 발생하는 공간의 특성을 연구하는 연구가 있고[4], 도시 내의 강력범죄의 발생 패턴을 사회적, 공간적, 경제적으로 분석한 후, 범죄가 발생할 유형적, 공간적, 시간적으로 예측하는 프로그램 개발의 필요성을 강조한 범죄 발생 예측 프로그램의 설계에 대한 연구가 진행되었다[5]. 또한 범죄자와 피해자의 거주지 사이의 공간분포를 조사하여, 범죄가 발생에 특정 패턴이 있음을 확인하고 패턴의 원인을 찾는 연구가 있었다[6].

2.2 마코프 체인

마코프 체인은 마코프 성질을 가진 이산 시간 확률 과정으로 러시아 수학자인 안드레이 마르코프의 이름에서 유래했다 [7]. 마코프 체인은 한 상태에서 다른 상태로 변환될 확률이 과거의 상태보다 현재의 상태에 의존하는 모델을 의미한다. 어떤 랜덤과정 $\{X(t)|t \geq 0\}$ 이 임의의 시점 $t_0 < t_1 < \dots < t_n$ 에 대해 $X(t_0), X(t_1), \dots, X(t_{n-1})$ 이 주어진 경우에 대한 $X(t_n)$ 의 조건부 누적분포함수가 $X(t_{n-1})$ 에만 의존한다면,

$$P[X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, \dots, X(t_0 = x_0)] = P[X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}] \tag{1}$$

식(1)로 기술할 수 있고 랜덤과정 $\{X(t)|t \geq 0\}$ 은 마코프 과정이다. 상태공간의 성질에 따라 연속상태 마코프 과정과 이산상태 마코프 과정으로 나누는데, 이산 상태의 마코프 과정을 마코프 체인이라고 한다[8]. 마코프 체인은 상태집합, 초기 확률, 상태 전이행렬로 구성되어 있다[9,10]. 마코프 체인은 시간에 따른 미래를 예측할 수 있는 것으로 다양한 분야에 적용되었다. 예를 들면, 공동주택에서 쾌적한 실내 환경을

유지하기 위한 재실자의 행동을 예측하는 연구에 사용 되었고 [11]. 전력 계통 시스템의 과거 고장데이터를 이용하여 다음에 발생할 전력기기의 고장 건수 예측에 마코프 체인을 적용하여 예측하였다[12].

III. 본 론

지금 까지 범죄 예측을 위해 적용된 마코프 체인 모델은 어떤 사건이 시작된 처음의 상태에서 시작하여, 예측하고 싶은 시간의 바로 전 시간의 상태까지의 상태인 전체 상태를 이용하여 확률 값을 구하였다[2,9,10]. 그러나 이 모델은 과거의 상태와 최근의 상태를 동등하게 처리하여, 최근에 변화 하는 상태들이 마코프 체인 모델의 결과에는 영향을 적게 주는 단점이 있다. 본 논문에서는 이런 점을 보완하기 위하여 범죄 예측 모델에 사용될 마코프 체인 확률 값을 전체 상태 마코프 체인 확률과 최근 상태 마코프 체인 확률로 나누어서 확률 값을 구하였다. 그리고 두 확률 값의 평균값을 범죄 발생 확률로 정의하였다. 범죄 예측 모델에 사용될 범죄 발생 확률을 구하는 모델은 그림 1과 같고 그 순서는 다음과 같다.

- ① 마코프 체인 모델에서 전체 상태를 이용하여 사건 발생 확률을 구한다.
- ② 마코프 체인 모델에서 최근의 상태를 이용하여 사건 발생 확률을 구한다.
- ③ 전체 상태의 확률 값과 최근 상태의 확률 값의 평균 확률 값을 범죄 발생 확률로 정의한다.

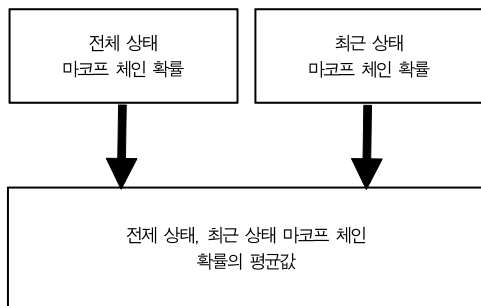


그림 1. 범죄 예측 확률 모델
Fig. 1. A crime prediction probability model

3.1 범죄 발생 예측 확률

매년 대검찰청에서 발표하는 범죄 통계 자료를 데이터화

하여 범죄 예측 확률 모델을 적용하여 범죄 발생 건수를 예측하였다. 우선 범죄 통계 자료를 바탕으로 임계값과 상태를 정의한다. 임계값과, 상태를 이용하여 상태 집합, 초기 확률, 전이행렬을 구한 후, 초기 확률과 전이 행렬을 이용하여 범죄 예측 확률 값을 계산한다[9,10].

◎ 상태 집합(S): 범죄 예측 확률 모델에서 상태는 한 범죄가 발생하는 빈도수의 범위이다. 범죄 발생 통계 자료를 바탕으로 임계값을 설정하였고, 임계값을 바탕으로 상태들을 집합으로 정의 하였다.

◎ 초기 확률: 하나의 범죄가 초기 상태에 가질 수 있는 범죄 발생 확률이다. 최근에 발생한 범죄 발생한 상태를 적용하여 식(2)과 같이 정의 한다.

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

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

여기서 a, b, c는 각 상태들(S_1, S_2, \dots, S_n)의 범죄 발생 건수이고 F는 a, b, c의 합이다. 또한, 확률이므로 초기 확률의 총합은 1인 식(3)을 만족한다.

◎ 전이 행렬: 각 범죄 발생 건수를 상태 집합에 매핑하여 상태들을 열거 한다. 그리고 열거된 상태들 사이의 전이 횟수를 구한 후, 전이 행렬로 나타낸다. 그리고 정의된 상태가 다른 상태로 전이하는 것을 확률로 나타낸 것이다. 식 (4)의 P는 전이 행렬로서 조건 식(5)을 만족한다. 단, 각 행의 합은 1을 만족 한다.

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

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

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

본 논문에서는 전이행렬은 두 가지로 나뉜다. 하나는 전 상태 상태를 가진 전이 행렬과, 최근 상태를 적용한 전이 행렬

이다. 전체 상태를 가진 전이 행렬과 최근 상태를 적용한 전이행렬을 이용하여 각각의 범죄 발생 확률 값을 구한 후 그것의 평균값을 계산한다.

◎ 범죄 예측 확률: 식(6)과 식(7)은 미래 예측 확률 식인 마코프 체인으로 정의 된다. 여기서 전체 상태를 가진 전이행렬(FP)과, 최근 상태를 적용한 전이행렬(RP)라고 하면

$$\text{전체 상태에 대한 예측 확률 (FP): } FP = \sum_{i=1}^n P(S_i) fP_{ik} \quad (6)$$

$$\text{최근 상태에 대한 예측 확률 (RP): } RP = \sum_{i=1}^n P(S_i) rP_{ik} \quad (7)$$

$P(S_i)$: 초기 확률

fP_{ik} : 전체 상태 전이 행렬, rP_{ik} : 최근 상태의 전이행렬

본 논문에서 제시한 적용할 범죄 예측 확률은 식(6)의 결과 값과 식(7)의 결과 값의 평균값을 적용한 식(8)로 한다.

$$\text{범죄 발생 예측 확률 (PP): } PP = \frac{FP+RP}{2} \quad (8)$$

◎ 범죄 발생 예측 건수 :

본 논문에서는 범죄 예측 건수를 예측 할 때 평균값과 최대값을 적용하여, 계산하고 비교한다.

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

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

n : 범죄 발생 상태 집합의 상태들의 수

$P(S_i)$: 범죄 발생 예측 확률

$M(S_i)$: 각 범죄 발생 기간에 발생한 범죄 발생 건수의 평균 값

(최근 5개월간 발생한 범죄의 평균 발생 건수)

$Max(S_i)$: 각 범죄 발생 기간에 발생한 범죄 발생 건수 중 최대값

3.2 범죄 예측 확률 모델링

본 논문에서는 앞 절에서 제시한 범죄 발생 예측 확률 모

델을 바탕으로 범죄 발생 예측 확률 모델링을 구성하였다. 범죄 발생 예측 확률 모델링에 범죄 발생 건수 데이터를 적용하여 범죄 발생 건수를 예측하였다. 범죄 예측 확률 모델링은 다음의 순서로 적용되며 그림 2는 범죄 예측 확률 모델링이다.

첫 번째, 월별 범죄 통계 자료를 수집하고 임계값 및 상태를 정의한다.

두 번째, 앞에서 정의한 임계값 및 상태를 이용하여 초기 확률을 계산한다.

세 번째, 정의된 상태 값을 이용하여 실험하려는 대상의 전체를 대상으로 하는 전체 상태 값과 최근 상태로 나눈다. 본 논문에서는 실험 데이터로 2005년~2008년의 범죄 발생 데이터를 이용했는데, 전체 상태는 2005년~2008년의 발생한 범죄 발생 데이터를 의미하고, 최근 상태는 가장 마지막 년도인 2008년에 발생한 범죄 발생 데이터로 정의하였다.

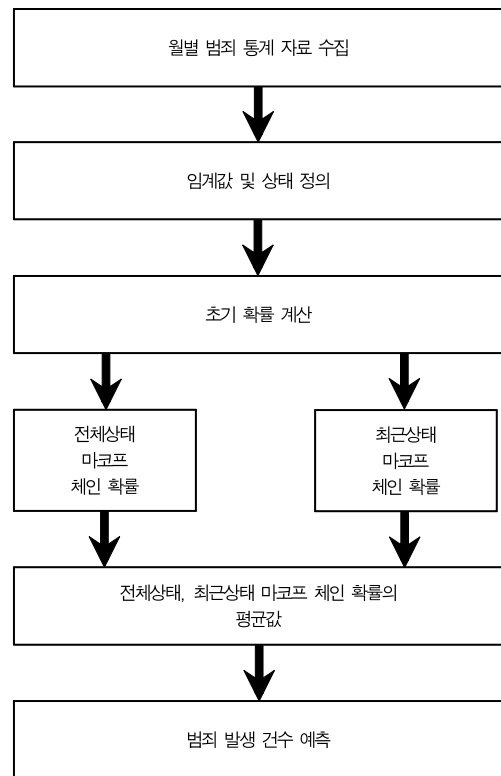


그림 2. 범죄 예측 확률 모델링
Fig. 2. A Crime probability of the prediction Modeling

네 번째, 전체 상태를 이용하여 마코프 체인 확률과 최근 상태를 이용하여 마코프 체인 확률을 계산한다.

다섯 번째, 전체 상태의 마코프 체인 확률과 최근 상태의 마코프 체인 확률의 평균값을 구한다.

마지막으로 각 범죄 발생 건수의 최대값과 평균값을 적용하여 범죄 발생 예측 건수를 계산한 후 비교한다.

IV. 범죄 발생 데이터를 범죄 예측 확률 모델링에 적용

다양한 범죄 중에서 과실 치사 범죄, 위조 범죄를 대상으로 범죄 발생 예측 건수를 구하였다.

4.1 과실범죄

과실범죄는 과실치사상, 업무상과실치사상, 실화가 포함된 범죄이다. 표1은 2005년부터 2008년까지의 과실범죄 발생 건수 이다. 과실범죄의 발생 건수를 분석하여 임계값과 상태를 정의 하였다.

표 1. 과실범죄발생건수
Table 1. The Incidence of criminal negligenc

| | 2005년 | 2006년 | 2007년 | 2008년 |
|-----|-------|-------|-------|-------|
| 1월 | 561 | 536 | 543 | 470 |
| 2월 | 409 | 505 | 474 | 405 |
| 3월 | 565 | 520 | 502 | 476 |
| 4월 | 553 | 449 | 488 | 511 |
| 5월 | 530 | 563 | 498 | 546 |
| 6월 | 516 | 486 | 481 | 454 |
| 7월 | 494 | 438 | 454 | 487 |
| 8월 | 519 | 456 | 491 | 406 |
| 9월 | 480 | 496 | 422 | 459 |
| 10월 | 535 | 520 | 440 | 511 |
| 11월 | 548 | 606 | 508 | 470 |
| 12월 | 543 | 537 | 474 | 557 |

◎범죄 발생 상태(S)

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

◎임계값 범위

$$S_1 : 0 \sim 450 \quad S_2 : 451 \sim 540 \quad S_3 : 541 \sim 630$$

◎초기 확률

2005년부터 2008년까지 발생한 과실 범죄 발생 중 최근 5개월에 발생한 범죄 건수에 식(2)을 적용하여 초기 확률 식(11)구하였다.

▶ 범죄 발생 건수

406, 459, 511, 470, 557

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

▶ 초기 확률 : $P(S_1:1, S_2:3, S_3:1)$

$$P(0.2, 0.6, 0.2) \quad (11)$$

4.1.1 전체 상태를 적용한 다음 달 과실 범죄 발생 확률
표1의 과실 범죄 발생 건수를 임계값 범위의 상태와 매핑하여 범죄 발생 상태를 나타낸다.

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

각 상태들(S_1, S_2, S_3)이 다른 상태로 전이하는 전이 확률을 식(4)을 이용하여 전이확률로 만든 후, 전이행렬로 표현하면 식 (12)로 구할 수 있다.

$$\begin{matrix}
 & S_1 & S_2 & S_3 \\
 S_1 & & & \\
 S_2 & & & \\
 S_3 & & &
 \end{matrix}
 \begin{pmatrix}
 0.14 & 0.57 & 0.29 \\
 0.16 & 0.68 & 0.16 \\
 0.11 & 0.67 & 0.22
 \end{pmatrix}
 \quad (12)$$

식(13)의 전체 상태 전이 다이어그램은 그림 3과 같다.

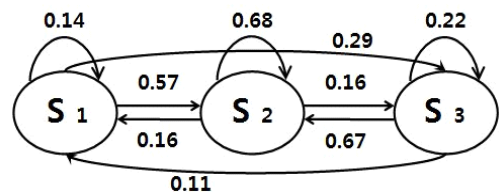


그림 3. 과실범죄에 대한 전체 상태 전이 다이어그램
Fig. 3. The Criminal negligence for the entire state transition diagram

초기 확률 식(11)과 전이행렬 식(12)을 구한 후 식(6)을 적용 하면 전체 상태에 대한 과실 범죄 발생 확률을 구할 수 있다.

$$(0.2 \ 0.6 \ 0.2) \begin{pmatrix} 0.14 & 0.57 & 0.29 \\ 0.16 & 0.68 & 0.16 \\ 0.11 & 0.67 & 0.22 \end{pmatrix} \quad (13)$$

$$= (0.15 \ 0.65 \ 0.2)$$

구해진 식(13)에 의해 과실범죄의 전체 상태에 대한 다음 달 발생 확률은 S_2 상태인 0.65일 때 가장 높다. 범죄 발생 예측 건수를 구하기 위해 범죄 발생건수의 평균값과 최대값을 식(9)과 식(10)에 적용하면 범죄 발생 예측 건수를 구할 수 있다.

▶평균값

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) M(S_i) \\ &= 0.65 \times 490 = 319 \end{aligned}$$

▶최대값

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) Max(S_i) \\ &= 0.65 \times 606 = 394 \end{aligned}$$

범죄 발생 건수를 예측한 결과 평균값을 사용하였을 때 319건, 최대값을 사용하였을 때 394건이 발생될 것으로 예측 되었고 실제 2009년 1월 달 과실 범죄 발생 건수는 421 건 이었다.

4.1.2 최근 상태를 적용한 다음 달 과실 범죄 발생 확률

표 1의 과실 범죄 발생 건수 중 최근 1년간(2008년)의 범죄발생건수를 매핑 하여 상태를 나타낸다.

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

각 상태들(S_1, S_2, S_3)이 다른 상태로 전이하는 전이 횟수를 식(4)을 이용하여 전이확률로 만든 후, 전이행렬로 표현하면 식 (14)로 구할 수 있다.

$$\begin{matrix} & S_1 & S_2 & S_3 \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \end{matrix} & \begin{pmatrix} 0 & 1 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 1 & 0 \end{pmatrix} \end{matrix} \quad (14)$$

식(14)의 상태 전이 다이어그램은 그림 4와 같다.

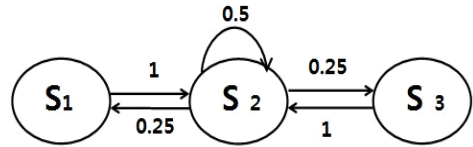


그림 4. 과실범죄에 대한 최근 상태 전이 다이어그램
Fig. 4. The Criminal negligence for the recent state transition diagram

초기 확률 식(11)과 전이행렬 식(14)을 구한 후 식(7)을 적용 하면 최근 상태에 대한 과실 범죄 발생 확률을 구할 수 있다.

$$(0.2 \ 0.6 \ 0.2) \begin{pmatrix} 0 & 1 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 1 & 0 \end{pmatrix} \quad (15)$$

$$= (0.15 \ 0.7 \ 0.15)$$

구해진 식(15)에 의해 과실범죄의 최근 상태에 대한 다음 달 발생 확률은 S_2 상태인 0.7이다. 전체 상태와 최근 상태의 발생확률은 식(8)에 의해 범죄 예측 확률(PP)은 0.68이다. 범죄 발생 건수의 평균값과 최대값을 이용하는 범죄 예측 건수 식(9)과 식(10)에 범죄 예측 확률을 적용하면 범죄 발생 예측 건수를 구할 수 있다.

▶평균값 : 마지막 5개월간의 범죄 발생 건수
평균 값

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) M(S_i) \\ &= 0.68 \times 490 = 333 \end{aligned}$$

▶최대값 : 2005년~2008년 사이에 발생한 범죄 발생건수 최대값

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) \text{Max}(S_i) \\ &= 0.68 \times 606 = 412 \end{aligned}$$

범죄 예측 확률(PP)를 이용하여 얻어진 확률로 범죄 발생 건수를 예측한 결과 평균값을 사용하였을 때 333건, 최대값을 사용하였을 때 412건이 발생될 것으로 예측 되었고 실제 2009년 1월 달 과실 범죄 발생 건수는 421 건 이었다.

4.2 위조범죄

위조범죄는 통화, 유가증권, 문서, 인장이 포함된 범죄이다. 표 2는 2005년부터 2008년까지의 위조범죄 발생 건수이다. 위조범죄의 발생 건수를 분석하여 임계값과 상태를 정의 하였다.

표 2. 위조범죄 발생건수
Table 2. Incidence of counterfeiting crimes

| | 2005년 | 2006년 | 2007년 | 2008년 |
|-----|-------|-------|-------|-------|
| 1월 | 2,021 | 2,686 | 1,992 | 2,120 |
| 2월 | 1,302 | 1,820 | 1,637 | 1,691 |
| 3월 | 2,185 | 1,919 | 2,073 | 1,801 |
| 4월 | 2,241 | 2,161 | 2,174 | 2,090 |
| 5월 | 2,036 | 3,013 | 2,678 | 2,025 |
| 6월 | 2,100 | 3,247 | 2,184 | 2,611 |
| 7월 | 1,808 | 3,492 | 2,228 | 2,755 |
| 8월 | 1,739 | 2,831 | 2,011 | 2,447 |
| 9월 | 1,836 | 2,038 | 1,703 | 3,372 |
| 10월 | 2,103 | 1,760 | 2,065 | 2,655 |
| 11월 | 2,370 | 1,898 | 2,464 | 2,062 |
| 12월 | 2,271 | 1,886 | 1,961 | 2,393 |

◎범죄 발생 상태(S)

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

◎임계값 범위

$$S_1: 0 \sim 2,100 \quad S_2: 2,101 \sim 2,900 \quad S_3: 2,901 \sim 3,700$$

◎초기 확률

2005년부터 2008년까지 발생한 위조범죄 발생 중 최근 5개월에 발생한 범죄 건수에 식(2)을 적용하여 초기 확률 식(16)구하였다.

▶ 범죄 발생 건수

$$2,447, 3,372, 2,655, 2,062, 2,393$$

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

▶ 초기 확률 : $P(S_1:1, S_2:3, S_3:1)$

$$P(0.2, 0.6, 0.2) \quad (16)$$

4.2.1 전체 상태를 적용한 다음 달 위조 범죄 발생 확률

앞의 절과 마찬가지로 표 2의 위조 범죄 발생 건수를 임계값 범위의 상태와 매핑하여 상태를 나타낸 후 각 상태가 다른 상태로 전이하는 전이 횟수를 식 (4)을 적용하여 전이 확률을 만든 후, 전이행렬로 표현하면 식(17)로 표현된다.

$$\begin{matrix} & S_1 & S_2 & S_3 \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \end{matrix} & \begin{pmatrix} 0.68 & 0.32 & 0 \\ 0.35 & 0.53 & 0.12 \\ 0.2 & 0.2 & 0.6 \end{pmatrix} \end{matrix} \quad (17)$$

초기 확률 식(16)과 전이행렬 식(17)을 구한 후 식(6)을 적용 하면 전체 상태에 대한 과실 범죄 발생 확률을 구할 수 있다.

$$\begin{aligned} (0.2 \ 0.6 \ 0.2) & \begin{pmatrix} 0.68 & 0.32 & 0 \\ 0.35 & 0.53 & 0.12 \\ 0.2 & 0.2 & 0.6 \end{pmatrix} \\ &= (0.39 \ 0.42 \ 0.19) \end{aligned} \quad (18)$$

구해진 식(18)에 의해 위조범죄의 전체 상태에 대한 다음 달 발생 확률은 S_2 상태인 0.42일 때 가장 높다. 범죄 발생 예측 건수를 구하기 위해 범죄 발생건수의 평균값과 최대값을 식(9)과 식(10)에 적용하면 범죄 발생 예측 건수를 구할 수 있다.

▶평균값

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) M(S_i) \\ &= 0.42 \times 2585.8 = 1086.04 \end{aligned}$$

▶최대값

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) \text{Max}(S_i) \\ &= 0.42 \times 3,492 = 1466.64 \end{aligned}$$

범죄 발생 건수를 예측한 결과 평균값을 사용하였을 때 약 1,086건, 최대값을 사용하였을 때 약 1,467건이 발생될 것으로 예측 되었고 실제 2009년 1월 달 과실 범죄 발생 건수

는 1,800건 이었다.

4.2.2 최근 상태를 적용한 다음 달 위조 범죄 발생 확률

표 2의 위조 범죄 발생 건수 중 최근 1년간(2008년)의 범죄발생건수를 매핑 하여 상태를 나타낸 후, 각 상태들이 다른 상태로 전이하는 전이 확률을 식(4)을 이용하여 전이확률로 만든 후, 전이행렬로 표현하면 식(19)으로 구할 수 있다.

$$\begin{matrix} S_1 & S_2 & S_3 \\ S_1 & \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.4 & 0.4 & 0.2 \\ 0 & 1 & 0 \end{pmatrix} \\ S_2 & \\ S_3 & \end{matrix} \quad (19)$$

초기 확률 식(16)과 전이행렬 식(19)을 구한 후 식(7)을 적용 하면 최근 상태에 대한 위조 범죄 발생 확률을 구할 수 있다.

$$(0.2 \ 0.6 \ 0.2) \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.4 & 0.4 & 0.2 \\ 0 & 1 & 0 \end{pmatrix} = (0.36 \ 0.52 \ 0.12) \quad (20)$$

구해진 식(20)에 의해 위조범죄의 최근 상태에 대한 다음 달 발생 확률은 S_2 상태인 0.52이다. 전체 상태와 최근 상태의 발생확률을 식(8)에 의해 범죄 예측 확률(PP)은 0.47이다. 범죄 발생 건수의 평균값과 최대값을 이용하는 범죄 예측 건수 식(9)과 식(10)에 적용하면 범죄 발생 예측 건수를 구할 수 있다.

▶평균값 : 마지막 5개월간 위조 범죄 발생 건수 평균 값

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) M(S_i) \\ &= 0.47 \times 2,585.8 = 1215.3 \end{aligned}$$

▶최대값: 2005년~2008년 사이에 발생한 위조 범죄 발생 건수 최대값

$$\begin{aligned} \text{범죄 발생 예측 건수} &= \sum_{i=1}^n P(S_i) Max(S_i) \\ &= 0.47 \times 3,492 = 1,641 \end{aligned}$$

범죄 예측 확률(PP)을 이용하여 얻어진 확률로 범죄 발생 건수를 예측한 결과 평균값을 사용하였을 때 약1,215건, 최대값을 사용하였을 때 약 1,641건이 발생될 것으로 예측되었고 실제 2009년 1월 달 위조 범죄 발생 건수는 1,800건 이었다.

4.3 예측된 범죄 발생 건수 분석

범죄 발생 예측 모델링에 과실범죄, 위조범죄를 적용하여 범죄 발생 건수를 구하였다. 표 3 은 2009년 1월 실제 발생한 범죄 발생건수와 각 예측 모델링으로 예측된 범죄 발생 건수를 비교한 것이다.

표 3 범죄 발생 건수와 예측된 범죄 발생 건수 비교
Table 3. Comparing the number of occurrences predicted the incidence of crime and the number of crimes

| | 2009년 1월 실제범죄발생건수 | 마코프체인 적용 예측 발생 건수 | | 범죄발생예측 모델링적용 예측발생건수 | |
|------|-------------------|-------------------|-------|---------------------|-------|
| | | 평균값 | 최대값 | 평균값 | 최대값 |
| 과실범죄 | 421 | 319 | 394 | 333 | 412 |
| 위조범죄 | 1,800 | 1,086 | 1,467 | 1,215 | 1,641 |

과실 범죄와 위조 범죄의 발생 건수가 실제 범죄 발생 건수와 어느 정도 유사한지 비교하였다. 각 범죄의 실제 발생 건수를 기준으로, 각 예측 모델링의 결과 값인 범죄 발생 예측 빈도수와 나누어 비교하면 그림 5 와 같다.

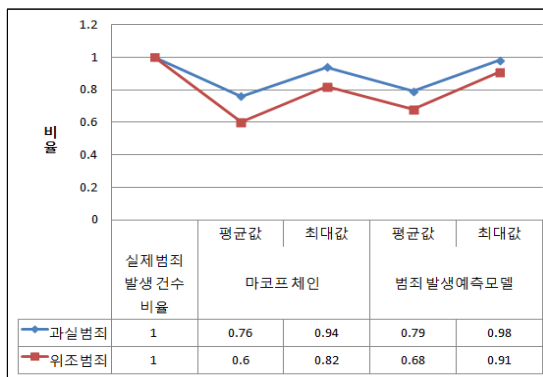


그림 5. 범죄 발생 건수와 예측된 범죄 발생 건수 비율
Fig. 5. Ratio of the number of occurrences predicted the incidence of crime and the number of crimes

그림 5에서와 같이 실제 범죄 발생 건수 비율을 1로 하였을 때, 과실 범죄의 경우 범죄 발생 예측 모델의 최대값을 적용하였을 때 0.98로 가장 높았다. 그리고 위조범죄의 경우에도 범죄 발생 예측모델의 최대값을 적용하였을 때, 0.91로 가장 높았다. 결론적으로 범죄예측모델링에 범죄 예측 모델에 적용된 기간 중 발생한 범죄 발생 건수 최대값을 적용하였을 때, 실제 범죄 발생 건수와 유사하였다.

V. 결 론

본 논문은 현대 사회의 큰 문제 중의 하나인 범죄를 예측을 위해 마코프 체인을 적용한 범죄 예측 모델링을 제안하였고, 데이터로는 과실범죄, 위조범죄를 적용하였다. 본 논문이 이전 범죄 예측 모델링 논문과 다른 점은 마코프 체인을 이용하여 미래 예측 확률을 구할 때, 전체 상태를 대상으로 하지 않고, 전체 상태와 최근 상태로 나누어 그 평균 확률 값을 적용한 것으로, 최근에 변화하는 범죄 상태를 범죄 예측 확률 값에 적용할 수 있었다. 그 결과 본 논문이 제시한 범죄 예측 모델링인 전체 상태와 최근 상태의 평균 확률 값을 적용하였을 때, 실제 범죄 발생 예측 건수가 유사하였다. 본 논문이 제시한 범죄 예측 모델링을 범죄 예방 정책에 적용한다면, 범죄 예방 활동이나 범죄 정책 수립에 도움이 될 것으로 예상된다. 그러나 싸이코 패스나 연쇄 살인범 같은 특수한 범죄는 예측하기 어렵고, 범죄 예측 모델링에 적용된 마코프 체인 특성상 가까운 미래만을 예측 할 수 있다. 이러한 점을 극복하기 위한 연구가 진행되어야 할 것이다.

참고문헌

[1] Building of Foresight knowledge base and Identification of future emerging technology, http://www.kistep.re.kr/policy/k2/k2_01_02.jsp

[2] 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, Vol 6, No 1, pp. 69-77, Feb. 2011.

[3] Chul-Hyun Park, "The Specialization of Criminal Career: A Review on the Techniques of Analysis on Specialization", Korean Criminological Review, Vol 73-3, pp. 999-1020, Sep. 2007.

[4] Sun-Young Hwang, Chul-sue Hwang, "The Spatial Pattern Analysis Of Urban Crimes Using GIS : The Case of Residential Burglary", Journal of the Korea Planners Association, Vol 38, No 1, pp. 53-66, Feb. 2003.

[5] Young-Hwan Kim, Jeong-Min Mun, "A Study on the Development of Crime Prediction Program", Journal of the Korea society of computer and information, Vol.11 No. 4, pp. 221-230 2006.

[6] Brown, MA., "Modelling the Spatial Distribution of Suburban Crime", Economy Geography, Vol. 58, No. 3, pp.247-261, 1982.

[7] Charles M. Grinstead, "Introduction to Probability: Second Revised Edition", American Mathematical Society, pp. 405-407, 1997.

[8] Oliver C. Ibe, "Fundamentals of Applied Probability and Random Processes", hantee media, pp. 371-393, 2008.

[9] Young-Gab Kim, Young-kyo Baik, 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, Aug. 2006.

[10] Won-Hyung Park, Young-Jin Kim, Dong-Hwi Lee, Kui-Nam J Kim, "A Study on Prediction of Mass SQL Injection Worm Propagation Using The Markov Chain", Journal of the Korea Institute of Information Security and Cryptology, Vol 8, No4, pp.174-181, Dec. 2008.

[11] Young-Jin Kim, Cheol-Soo Park, "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.

[12] Hee Tae-Lee, Jae-Chul Kim, "A Study on The Prediction of Number of Failures using Markov Chain and Fault Data", KIIEE Annual Autumn Conference 2008, pp. 363-366, Oct .2008.

저 자 소 개



박 구 락

1986: 중앙대학교 전기공학과 공학사.

1988: 숭실대학교

전자계산학과 공학석사.

2000: 경기대학교

전자계산학과 이학박사.

현 재: 공주대학교 컴퓨터공학부 교수

관심분야: 정보경영, 정보통신,

전자상거래

Email : ecgrpark@kongju.ac.kr