한국청소년연구
Studies on Korean Youth
2015. Vol. 26. No. 2. pp. 35~57.
ISSN 1225-6336(Print) / 2288-8799(Online)
http://dx.doi.org/10.14816/sky.2015.26.2.35

마코프 체인 프로세스에 기반한 학교폭력 발생 예측 모형

손 재 환*

초 록

본 연구는 최근 사회적으로 이슈화된 학교폭력 문제를 예측할 수 있는 확률 모형을 제안하고자 했다. 이를 위해 공학, 정보학 등 여러 분야에서 미래 예측을 위해 사용하고 있는 마코프 체인 프로세스 확률 모형을 사용하였다. 이에 사용된 분석 자료는 최근 3년간(2012. 6. ~ 2014. 6) 117학교폭력신고센터에 접수된 신고 건수로서 월별 자료를 분석하였다. 학교폭력 발생 빈도에 대한 분포 양상을 분석한 뒤 빈도에 대한 상태를 관심(S1), 주의(S2), 경계(S3)로 정의하여 분류한 뒤, 상태 전이 행렬을 산출하였다. 또한 최근 5개월간의 학교폭력신고건수 빈도를 통해 초기 확률을 산출하고 이 값과 전이행렬을 곱하여 다음 달 예측 가능한 학교폭력 상태 및 빈도를 산출하였다. 결과로는 2014년 7월 학교폭력 상태가 주의단계로 나타났으며, 예측빈도가 7,565.86건으로 나타나 실제 관찰된 7,422건과 유사한 결과가 나타났다. 또한 마코프 체인 프로세스를 통해 예측된 2014년 7월의 수치를 다른 예측모형을 통해 나타난 수치와 비교하였다. 비교된 예측모형은 회귀모형과 시계열모형이었다. 우선 학교폭력 건수를 종속변인으로 각 월을 독립변인으로 설정한 회귀모형에서는 예측된 2014년 7월의 수치가 6,970.76건으로 나타났다. 그리고 시계열분석은 대표적으로 가장 많이 사용되는 ARIMA 모형을 사용하였으며 이를 통해 예측된 2014년 7월의 수치는 8,002.60건이었다. 따라서 각 예측모형에서 산출된 예측치를 비교해 볼 때 마코프 체인 프로세스 모형이 실제 관찰된 값과 가장 유사한 것으로 나타났다.

주제어: 학교폭력, 마코프 체인 프로세스, 예측 모형

* 건양대학교, kiki5048@hanmail.net

Ⅰ. 서 론

최근 지난 3년간 학교폭력 문제는 우리사회 청소년 문제 중 가장 이슈화된 문제로 국민들에게 알려졌다. 2012년 학교폭력 피해로 인한 청소년 자살 등 심각한 문제들이 언론을 통해 알려지자 학교폭력에 대한 심각성이 국가적 차원의 문제로 떠오르게 되었다.

학교폭력에 대한 정의는 그 유형과 형태가 너무나 다양해서 학자들 마다 다르게 정의하고 있다. 2012년에 개정된 학교폭력예방 및 대책에 관한 법률에 따르면 학교폭력이란 '학교 내외에서 학생을 대상으로 발생한 상해, 폭행, 감금, 협박, 약취·유인, 명예훼손·모욕, 공갈, 강요·강제적인 심부름 및 성폭력, 따돌림, 사이버 따돌림, 정보통신망을 이용한 음란·폭력 정보 등에 의하여 신체·정신 또는 재산상의 피해를 수반하는 행위'로 정의되고 있다(손재환 외, 2012). 이 정의에 따르면 학교폭력의 유형과 형태가 매우 폭이 넓은 것을 알 수 있다. 하지만 이 외에도 학교폭력의 형태는 해가 지나면서 이전과 다른 새로운 형태가 나타나고 있다. 빵을 사오라고 심부름을 시키는 소위 빵서틀, 한 아이의 스마트폰을 강제로 공유해서 인터넷을 사용하는 와이파이셔틀, 스마트폰 등 매체를 이용한 SNS 폭력 등은 갈수록 학교폭력의 형태가 우리 청소년들에게 다양화되고 있음을 반영하고 있다. 또한 학교 폭력의 심각성도 갈수록 증가하고 있다. 같은 반 여 학우를 강제로 유인하고 협박하여 성매매를 시키고 살해하거나 절도 등 범죄 수단에 강제로 동참하게 하는 등의 심각한 범죄에서부터 지속된 언어 폭력과 희롱 등 심각한 수준의 심리적 외상을 유발하게 하는 등 는 등 매우 파괴적인 폭력 형태들이 각종 언론에 보도되고 있다. 국회 안전행정위원회 조원진(2014) 의원의 보도자료에 따르면 경찰청에 보고된 학교폭력 검거자는 지난 5년간(2010년~2014년) 무려 95,739명이다. 이 자료에 따르면 매년 약 2만명 수준의 학교폭력 가해자가 검거되고 있는 실정이다. 같은 보도자료에 따르면 지난 5년간 가장 많이 검거된 학교폭력 유형으로는 폭력이 60,782명(63.5%), 금품갈취 19,359명(20.2%) 순이었으며 학교별로는 중학교가 가장 많았고(45.3%) 그 다음으로 고등학교(36.9%)로 많았다. 또한 검거된 초등학생도 3.1%나 차지하여 학교폭력이 저연령대에서 심각한 문제임을 나타내고 있다.

학교폭력 문제가 점차 심각해지자 정부는 '학교폭력 근절 종합대책'을 수립하여 교

육과학기술부, 경찰청, 여성가족부 등 범부처 차원의 대응 방안을 마련하여 117 학교폭력 신고센터, Wee센터, 청소년상담복지센터 등 전문기관을 통해 예방 및 개입 서비스를 제공하고 있다. 특히 117 학교폭력신고센터는 학교폭력과 관련된 초기 상담접수의 역할을 맡고 있다. 정부는 학교폭력을 예방하고 이에 적극 대응하고자 117센터에 대한 적극적인 홍보를 전개하고 전화 또는 인터넷 등을 통해 신고접수 및 상담을 실시하고 있으며 문제가 심각할 경우 심리 상담을 위한 전문기관 의뢰, 경찰 등의 사법적 의뢰 등을 진행하고 있다. 조원진(2014)에 따르면 117 학교폭력신고센터가 개소한 2012년 6월 이후 학교폭력 신고건수는 하루 평균 267건에 달하며 총 21만 3천여 건이 신고 접수된 것으로 나타났다. 이중 본인이 신고한 건수는 14만 7천여 건(69.1%)이며 나머지는 타인에 의해 접수 되는 것으로 나타났다. 타인에 의한 접수가 약 30%에 달하는 것은 학생들이 보복 폭행이 두려워 신고를 망설이고 있다는 것으로 추측될 수 있다.

이처럼 학교폭력 문제는 정부에서 적극 개입할 정도로 국가적 차원의 문제이기 때문에 이를 사전에 예방할 수 있는 노력이 필요하다. 학교폭력을 포함한 여러 문제들이 청소년들에게 발생하면 아직 어린 나이기 때문에 그 결과의 부정적 영향을 매우 크다. 따라서 문제 발생 후 개입 보다는 사전에 문제를 막을 수 있는 예방 노력이 훨씬 필요하다. 이를 위해서는 학교폭력에 대한 예방 교육 등 학생들의 인성 함양 프로그램도 필요하며, 문제 발생을 사전에 예측하고 이에 대응할 수 있는 예측 체계가 요구된다.

특정 문제 발생에 대한 예측 체계 구축의 필요성은 최근 빅데이터 연구를 통해 정부 및 공공기관, 일반 기업들에게 많이 인식되고 있다. 빅데이터 연구는 특정 현상의 비정형 혹은 정형화된 자료를 통해 자료의 패턴을 분석하고 미래의 현상을 예측할 수 있는 다양한 모형을 개발하고 있다. 빅데이터 분석 대상은 너무나 다양하며 범죄 발생 분석, 인터넷 바이러스 출현 분석, 독감 전염 경로 분석 등 미래의 현상을 예측하고 이를 대비하고자 하는 대상에 모두 적용될 수 있다. 일기예보 또한 방대한 기상자료라는 빅데이터 분석을 통해 미래를 예측하는 매우 훌륭한 예측시스템이라고 할 수 있다. 학교폭력을 포함한 청소년 문제도 이전 축적된 자료를 통해 발생 정도를 예측할 수 있다면 정부 혹은 관련기관에서 이에 대한 대비와 계획을 수립하는데 많은 도움이 될 것으로 예상되며 또한 청소년 문제를 예방하는데 있어 많은 도움이 될 것

으로 생각된다. 청소년 문제에 대한 예측 분석은 문제가 발생 된 후 감당해야할 사회적 위험을 사전에 감소시킬 수 있는 장점을 가진다.

따라서 본 연구에서는 학교폭력 발생을 사전에 예측할 수 있는 확률 모형을 제안하고자 한다. 이를 위해 마코프 체인 프로세스(markov chain process)라는 확률 모형을 사용하였다. 마코프 체인 프로세스는 특정 현상의 현재 상태를 파악함으로서 미래의 상태를 추측하는 확률 모형으로서 러시아 수학자 Andrey Markov에 의해서 개발된 확률 모형이다. 마코프 체인 프로세스는 과거 발생 데이터를 근거로 해서 시간의 흐름에 따라 발생 정도를 예측할 수 있는 모형이다. 이 모형은 다양한 분야에 사용되는데 인공지능 개발, 로봇의 움직임 개발, 일기예보, 범죄발생 예측, 웜 바이러스 발생 예측 등 현상과 행동을 예측하고자 하는 폭넓은 분야에 적용되고 있다. 일례로 노찬숙과 김동현(2012)은 2년간 범죄 발생 빈도 자료에 마코프 체인 프로세스를 적용하여 범죄 발생을 예측하는 확률 모형을 제시하였으며, 정영석과 정진영(2012)도 범죄 발생 예측 방법으로 마코프 프로세스를 적용하였다. 그리고 김영갑, 백영교, 인호와 백두권(2006)은 5년간 웜 바이러스 발생 빈도 자료를 마코프 체인 프로세스로 예측 가능한 모형을 제안하고 이에 대한 대비책을 시사하였다.

흔히 사회과학에서 예측 모형으로 자주 사용되는 회귀모형(regression model)은 변인 간 관계로부터 얻어진 회귀식을 통해 미래의 값을 예측할 수 있지만, 사용되는 변인 분포의 정상성을 가정하고 있어 관찰된 사례가 부족하거나 추세가 일정한 패턴을 보이지 않을 경우 정확한 예측이 어려운 단점을 가진다. 반면 마코프 체인 프로세스는 적은 사례의 이산(discrete) 자료 분석에 효과적이며 특히 시계열을 가진 자료 분석에 매우 유용하게 사용되고 있다.

일례로 범죄 발생이나 웜바이러스 출현 빈도는 여러 사회적 요인이나 환경 변화의 요인에 따라 매우 불특정한 빈도를 보인다. 그래서 단순 과거 빈도의 추세를 통해서는 미래를 예측하기가 어렵다. 마찬가지로 학교폭력을 비롯한 청소년문제의 경우에도 여러 사회·환경적 요인에 따라 발생빈도가 매우 불특정한 패턴을 보이는 경우가 많다. 청소년의 경우 독립성을 갖지 못한 시기이기 때문에, 경우 정치, 경제, 교육, 문화 등 사회 전반의 변화에 대해 영향을 받을 수 있는 여지가 많다. 하지만 이러한 모든 영향을 고려하여 특정 문제에 대한 미래를 예측하는 것은 매우 어려운 일일 것이다.

마코프 프로세스는 이러한 과거추세의 시계열 패턴이 없는 불특정 패턴을 가진 자

료 예측에도 유용하게 사용되고 있다. 일례로 동전을 던져 앞면이 나오는 패턴은 이전 시행에서 뒷면이 나왔다하더라도 이번 시행에서는 앞면이 나올 확률은 여전히 0.5이다. 즉 과거 추세의 영향을 받지 않는다는 것인데, 마코프 프로세스는 이러한 불특정 패턴의 자료에도 확률을 예측할 수 있는 모형으로 알려져 있다. 따라서 학교폭력 발생 빈도와 같이 불특정한 추세를 가진 빈도 자료를 예측하기에는 마코프 프로세스 모형이 적절한 것으로 사료된다.

따라서 본 연구에서는 노찬숙, 김동현(2012)과 김영갑 등(2006)의 연구를 참조하여 학교폭력 발생 빈도 자료에 대한 마코프 체인 프로세스를 적용하고자 했다. 이를 통해 미래 학교폭력 발생 정도를 예측하고자 했다. 이에 사용된 학교폭력 발생 빈도 자료는 현재 학교폭력에 대응하고 있는 가장 공신력 있는 기관인 117 학교폭력신고센터의 최근 3년간 월별 신고 접수 현황 자료를 사용하였다. 이 자료에 마코프 체인 프로세스 모형을 적용하여 미래 발생 가능한 학교폭력 발생 확률을 산출하고 실제 발생 빈도와 이를 비교하였다. 또한 자료의 예측을 위해 흔히 사용되고 있는 회귀분석과 시계열분석의 결과와 비교하여 예측의 정확성 정도를 비교하였다.

이를 통해 본 연구는 학교폭력 발생에 관한 예측 체계를 마련하는데 하나의 확률 모형을 제안하고자 하였으며 나아가 학교폭력 외에 다양한 청소년 문제를 예측하는데 도움이 되는 모형을 제안하고자 했다.

Ⅱ. 방 법

1. 마코프 체인 프로세스

본 연구에서는 학교폭력 발생 건수를 예측하고자 마코프 체인 프로세스에 기반한 모형을 사용하였다. 마코프 체인은 러시아의 수학자 Markov에 의해서 개발된 확률 모형으로 상태 간 전이가 오로지 이전 n개의 상태에 의존하여 이루어지는 프로세스를 말한다(김영갑 등, 2006). 여기서 n은 다음 상태를 결정하는데 영향을 미치는 상태의 개수를 말하며 전이는 상태의 변화를 말한다. 상태간의 전이는 이전의 상태에 의존하

여 이루어지고, 다음 상태를 결정하는 확률모델로서 현재의 시스템에 영향을 주고 그 전 단계의 상태에는 전혀 영향을 받지 않는 확률과정을 가정한 것으로 미래에 있을 변화를 예측하기 위한 기법이다(노찬숙, 김동현, 2012). 즉 마코프 체인 프로세스에서는 과거와 현재 상태가 주어졌을 때 미래 상태의 조건 확률 분포가 과거 상태와는 독립적으로 현재 상태에 의해서만 결정된다고 가정하는 확률 이론이다.

김영갑 등(2006)이 소개한 마코프 체인 프로세스를 자세히 살펴보면, $X(t)$가 확률과정일 때 $X(t)$가 취할 수 있는 값을 상태라 하고 이산(discrete) 값일 경우 $t_1 < t_2 < t_3 \cdots < t_n < t_{n+1}$ 에 대하여

$$P = P[X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, \cdots, X(t_1) = x_1] \quad (1)$$
$$= P[X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n]$$

로 마코프 성질이 기술된다. 식 (1)에서 $t_1, \cdots, t_{n-1}$은 과거시점이고, t_n은 현재, t_{n+1}은 미래이다. 마코프 프로세스의 상태가 이산 값이면 마코프 체인(markov chain)이라고 한다. 본 연구에서 분석되는 자료는 학교폭력 빈도이기 때문에 마코프 체인 프로세스를 사용하였다.

정리하자면 마코프 체인 프로세스는 과거에 있었던 변화를 근간으로 하여 시스템의 동적 성격을 파악하고 미래에 있을 변화를 예측하는 시스템을 모형화하는 수학적 기법으로 본 연구에서는 학교폭력 발생 빈도의 과거 변화에 대해 마코프 체인 프로세스를 통해 미래의 발생 정도를 예측하고자 하였다.

2. 마코프 체인 프로세스 진행 절차

본 연구에서는 김영갑 등(2006)이 제안한 마코프 프로세스 모델의 단계를 적용하여 학교폭력 발생 예측 모델을 제안하고자 하였다. 이는 크게 4단계에 걸쳐 진행되는 단계이며 진행 절차는 그림 1과 같다.

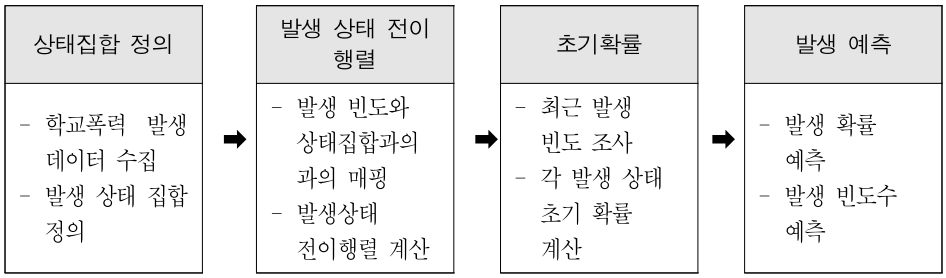


그림 1. 마코프 체인 프로세스 절차 도식

첫째, 상태집합의 정의 단계에서는 학교폭력 발생 빈도의 상태(state)를 정의한다. 여기서 상태는 학교폭력 발생 빈도에 따른 위험의 정도로 정의되며, 상태집합은 학교폭력 빈도의 위험 상태가 가질 수 있는 값들의 범위를 나타낸다. 본 연구에서 사용된 학교폭력 발생 빈도는 조원진(2014) 국회의원이 언론에 보도한 경찰청 2014년 7월 기준 학교폭력 현황 자료를 사용하였다. 이 자료에는 2012년 6월부터 2014년 7월 까지 학교폭력신고센터에 접수된 신고 건수가 나타나 있다.

표 1

학교폭력신고센터 월별 학교폭력 신고건수(조원진, 2014)

구분	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
'12년						10,923	9,927	5,134	9,651	10,800	10,736	8,838
'13년	4,730	6,033	10,575	12,203	12,026	10,896	8,876	4,383	6,686	8,561	8,577	7,978
'14년	3,082	3,910	7,184	8,577	8,212	7,580	7,422					

둘째, 발생 상태 전이 행렬은 상태집합에서 정의된 위험 상태와 발생 빈도 자료를 이용하여 위험 상태들 간의 전이 행렬을 구하는 단계이다. 분석된 각 위험별 발생 빈도와 전 단계에서 정의된 상태 집합과의 매핑(mapping)을 통해 학교폭력 발생 상태 전이행렬을 생성할 수 있다. 발생 상태 전이 행렬단계에서 생성되는 상태 전이 행렬은 식 (2)와 같으며 조건 (3)을 만족한다.

$$P=\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & P_{ij} & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix} \quad (2)$$

$$P_{ij} \geq 0,\ \sum_{j=1}^{n} P_{ij} = 1 \quad i = 1, 2, \cdots n \quad (3)$$

식 (2)에서 P_{ij}는 각 발생 상태에서 다른 발생 상태로의 확률로 이전 단계에서 정의된 각 상태집합의 전이 횟수를 전체 전이 횟수로 나눈 값이다. 예를 들어 각 빈도에 따른 상태집합이 저(S_1) - 저(S_1) - 고(S_3) 로 전이될 때, S_1에서 S_1으로 전이되는 횟수는 1개, S_1에서 S_2로 전이되는 횟수는 0개, S_1에서 S_3로 전이되는 횟수는 1, 전체 전이 횟수는 2이다. 따라서 S_1에서 S_1으로 전이되는 확률 값은 0.5, S_1에서 S_2으로 전이되는 확률 값은 0, S_1에서 S_3로 전이되는 확률 값은 0.5이다. 따라서 이에 대한 행렬식은 [0.5, 0, 0.5]로 나타낼 수 있다. 식 (3)은 모든 발생 상태에 대한 확률 값은 0보다 크거나 같아야 하며, 확률 값을 각 행의 합은 1이 되어야 한다는 조건을 나타낸다.

셋째, 초기 확률에서는 정의된 각 위험 상태가 초기 상태에 발생할 수 있는 확률을 구한다. 초기 확률 값을 구하기 위한 산식은 식 (4)와 같으며 조건 (5)를 만족한다.

$$P(S_1, S_2, \cdots, S_n) = P(\frac{a}{F}, \frac{b}{F}, \cdots, \frac{c}{F}) \quad (4)$$

$$F=\sum_{i=1}^{n} f_i = a + b + \cdots + c \quad (5)$$

$$\text{단 } f_i (i = 1, \cdots n) \text{ 이며 } \sum_{i=1}^{n} P(S_i) = 1 \quad (6)$$

식 (4)에서 a, b, c 는 최근 발생 빈도에서 정의된 각 상태집합의 횟수 이며, F는 각 상태집합의 횟수를 모두 더한 전체 상태집합 횟수이다(식 5). 즉 최근 자료에서 나타난 각 상태집합 횟수를 전체 상태집합 횟수로 나누어 준 것이 초기 확률 값이 된다. 이때 최근 발생 빈도에서 산출된 초기 확률 값들은 모두 더하면 1이 되어야 한다(식 6). 본 연구에서는 표 1의 학교폭력 발생 빈도 자료에서 2014년 2월 ~ 2014년 6월까지의 5개월 자료를 가지고 초기 확률 값을 산출하였다.

마지막, 발생 예측 단계에서는 전 단계에서 구한 위험 상태 전이행렬과 초기 확률 값을 통해 앞으로 발생할 위험 발생 확률이나 빈도를 예측한다. 아래 식 (7)은 식(2)와 식(4)를 이용한 학교폭력 발생 확률식이다.

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

식(7)에서 n은 학교 발생 상태 집합의 개수이고 $P(S_i)$는 초기 확률이며, P_{ik}는 전이행렬이다.

Ⅲ. 결 과

1. 상태집합 정의

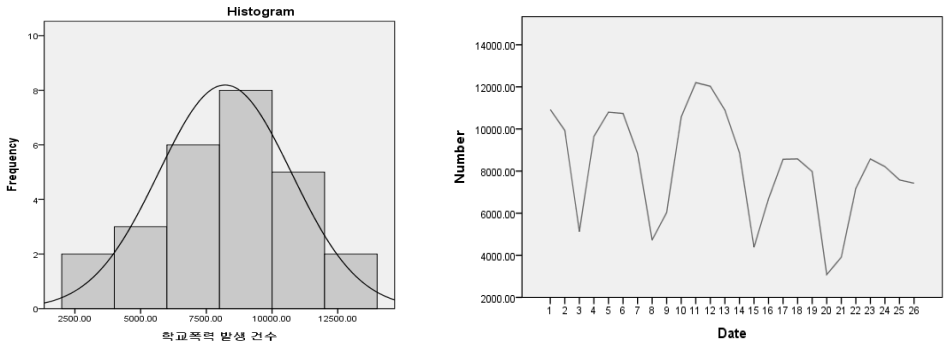
상태집합의 범위를 발생 빈도를 분석하여 위험 수준을 고, 중, 저로 구분하여 정의하였다. 고, 중, 저 구분을 위한 임계치는 발생 빈도 분포 양상을 고려하여 구분하였다. 빈도에 대한 분포 양상을 분석한 결과 정상분포 양상을 가지고 있었다(그림 2 a: 평균=8211.54, 표준편차=2531.47, 범위 3082~12203, n=26). 이에 평균으로 부터 표준편차 ±1을 임계치를 설정하여 상태집단을 세 가지로 구분하였다. 표준편차 -1 미만에 해당하는 빈도들의 경우 관심 상태(S_1), -1과 +1 사이에 있는 빈도들은 주의상태(S_2), 표준편차 +1 초과는 경계 상태 (S_3)로 정의하였다. 상태정의를 표준편차에 따라 집단으로 구분한 것은 발생 빈도 수준에 따른 학교폭력 상태에 대한 학문적 정의가 이루어지지 않은 상황에서 이를 구분하는 가장 적절한 기준은 빈도분포 양상에 따라 구분하는 것으로 절단 점을 설정하는 기준으로 가장 많이 사용되는 표준편차 값으로 정하고자 한 것이다. 또한 3 집단의 구분은 김영갑 등(2006), 노찬숙과 김동현(2012)의 연구의 권유에 따라 가장 단순한 형태인 3집단으로 구분 한 것인데, 이는 집단의 구분이 많으면 산출되는 행렬식이 복잡해지고 그에 따라 빈도가 없는 셀(null cell)이 생김으로 인해 최종 산출되는 확률 값이 부정확해질 수도 있기 때문이다.

상태집합에 대한 정의를 표 2에 정리하였다. 학교폭력 발생 상태 집합은 $S = \{S_1, S_2, S_3\}$ 로 표시된다.

표 2

학교폭력 발생 건수에 대한 상태집합 정의

상태 집단	S_1 : 관심 상태	S_2 : 주의 상태	S_3 : 경계 상태
발생 빈도 범위	5,680 미만	5,680 이상~10,743 미만	10,743 이상



(a) (b)

그림 2. 학교폭력 발생 빈도에 그래프 : a는 학교폭력 발생 빈도에 대한 히스토그램으로 정상분포하고 있었으며 b는 각 월별 시간에 대한 빈도의 추이를 그래프로 표시한 것이다.

2. 전이 행렬 산출

2012년 6월부터 2014년 6월까지의 빈도를 위에서 정의된 상태집합 S와 맵핑하여 상태를 아래와 같이 열거하였다. 표 3은 학교폭력 발생 빈도에 따른 상태집합 매핑을 나타내었다.

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

표 3

학교폭력 발생 빈도와 상태집합 매핑

구분	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
'12년						10,923	9,927	5,134	9,651	10,800	10,736	8,838
						S_3	S_2	S_1	S_2	S_3	S_2	S_2
'13년	4,730	6,033	10,575	12,203	12,026	10,896	8,876	4,383	6,686	8,561	8,577	7,978
	S_1	S_2	S_2	S_3	S_3	S_3	S_2	S_1	S_2	S_2	S_2	S_2
14년	3,082	3,910	7,184	8,577	8,212	7,580						
	S_1	S_1	S_2	S_2	S_2	S_2						

위 열거된 상태들로부터 각 상태에서 다른 상태로 전이되는 횟수를 구하고 이를 바탕으로 상태 전이 행렬식을 식(8)과 같이 구하였다.

$$\begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \end{array}\begin{array}{c} \begin{array}{ccc} S_1 & S_2 & S_3 \end{array} \\ \begin{bmatrix} 1 & 4 & 0 \\ 4 & 8 & 2 \\ 0 & 3 & 2 \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \end{array}\begin{array}{c} \begin{array}{ccc} S_1 & S_2 & S_3 \end{array} \\ \begin{bmatrix} 0.20 & 0.80 & 0.00 \\ 0.29 & 0.57 & 0.14 \\ 0.00 & 0.60 & 0.40 \end{bmatrix} \end{array} \tag{8}$$

식 (8)을 상태 다이어그램으로 나타내면 그림 3과 같다.

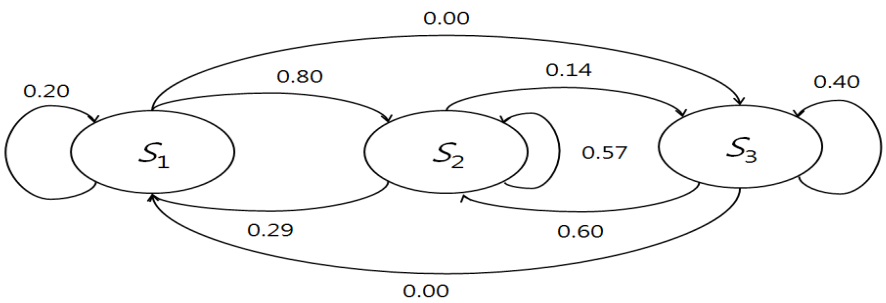


그림 3. 학교폭력 발생 상태 전이 행렬에 대한 다이어그램

3. 초기 확률 산출

학교폭력 발생 예측 모델에 적용하기 위한 초기 확률을 산출하기 위해서 최근 5개월간의 학교폭력 데이터를 사용하였다. 표 4에서 2014년 2월부터 2014년 6월까지의 상태집합을 살펴보면 S_1 1개, S_2 4개, S_3 0개 이다. 따라서 식(4)에 의하여 산출되는 초기 확률 값은 아래 식 (9)와 같다.

$$P = (0.2,\ 0.8,\ \ 0.0) \tag{9}$$

표 4

최근 5개월 간 학교폭력 발생 빈도와 상태집합 매핑

구분	'14년 2월	'14년 3월	'14년 4월	'14년 5월	'14년 6월
빈도	3,910	7,184	8,577	8,212	7,580
상태집합	S_1	S_2	S_2	S_2	S_2

4. 발생 예측

학교폭력 발생 상태전이 행렬식(8)과 초기확률식(9)를 이용하여 다음 월에 발생하게 될 학교폭력 발생 확률을 예측하고 또한 발생 빈도수를 예측할 수 있다. 이는 식 (7)에 의하여 식(10)과 같이 산출되었다.

$$P(S_k) = \sum_{i=1}^{n} P(S_i) P_{ik} = [0.2\ \ 0.8\ \ 0.0] \begin{bmatrix} 0.20 & 0.80 & 0.00 \\ 0.29 & 0.57 & 0.14 \\ 0.00 & 0.60 & 0.40 \end{bmatrix} = [0.27\ \ 0.62\ \ 0.11] \tag{10}$$

식 (10)으로부터 '14년 7월의 경우 S_1(관심 상태)는 0.27, S_2(주의 상태)는 0.62, S_3(경계 상태)는 0.11의 확률로 발생할 것이라고 예측된다. 즉 학교폭력 발생 건수가 5,680 이상~10,743 미만 일 것으로 가장 예측된다.

보다 구체적으로 '14년 7월에 예측 가능한 학교폭력 발생 빈도를 산출하기 위해서

아래 식 (11)과 (12)를 이용하였다.

$$\text{학교폭력 발생 예상 빈도} = \sum_{i=1}^{n} P(S_i)M(S_i) \quad (11)$$

$$\text{학교폭력 발생 예상 빈도} = \sum_{i=1}^{n} P(S_i)Max(S_i) \quad (12)$$

식 (11)과 (12)는 노찬숙과 김동현(2012)이 산출한 방식으로서 식 (11)의 경우 전체 빈도의 평균치(M = 8,211.54)와 다음 달 발생 확률 0.62를 곱하여 산출하는 식이며, 식 (12)는 빈도 중 최대 값(12,203)을 사용하여 예상 빈도를 산출하는 방식이다. 우선 (11)로 산출된 값은 5,091.15건이다. 그리고 (12)로 산출된 값은 7,565.86건이다. 따라서 식 (11)과 식(12)를 통해 얻을 수 있는 '14년 7월에 발생하게 될 학교폭력 발생 빈도는 5,091.15~7,565.86의 범위 내에서 발생할 것으로 예측된다. 실제 '14년 7월의 학교폭력 발생 건수는 7,422건으로 위 범위 내에 위치하는 것으로 확인되었다. 또한 최대 값을 통해 발생 빈도를 예측하는 것이 근사한 예측치를 산출하는 것으로 확인되었다. 이는 노찬숙과 김동현(2012)의 연구결과와 유사한 결과로 이들의 범죄발생 예측 연구에서도 최대치를 활용한 예측치가 관찰된 값에 보다 근사하게 나타났다.

- 식 (11)로 산출된 예상 빈도 : $\sum_{i=1}^{n} P(S_i)M(S_i) = 0.62 \times 8{,}211.54 = 5{,}091.15$
- 식 (12)로 산출된 예상 빈도 : $\sum_{i=1}^{n} P(S_i)Max(S_i) = 0.62 \times 12{,}203.00 = 7{,}565.86$

5. 다른 예측치와의 비교

위에 산출된 결과의 정확성을 보다 정밀하게 살펴보기 위해서 예측 방법으로 흔히 사용되는 회귀분석과 시계열분석방법으로 산출된 예측치와 비교하였다. 회귀분석은 주어진 자료에 대한 회귀식 산출을 통해 미래의 값을 예측하기 위해 흔히 사용되며, 시계열분석은 주식변동, 기업매출, 고객방문 수 등 값의 변동 폭이 큰 자료의 미래 값을 예측하는데 사용된다.

1) 회귀분석

우선 회귀분석을 통해 2014년 7월의 값을 산출하고자 하였다. 2012년 6월부터 2014년 6월까지의 빈도 값을 종속변인으로하고 월(month)을 독립변인으로 단순회귀분석을 실시하였다. 회귀분석으로 산출된 회귀식은 식(13)과 같다. 전체 설명변량은 R^2 = 0.111이었다.

$$\hat{Y} = -103.99x + 9674.5 \quad (13)$$

식 (13)의 x는 매 월의 수로 2014년 7월에 해당하는 x 값은 26이다. 이 값을 대입하면 $\hat{Y} = (-103.99 \times 26) + 9,674.5 = 6,970.76$ 이다. '14년 7월의 실제 학교폭력 발생 건수는 7,422건인데 이 값과 비교해 보면 마코프체인 프로세스를 통해 산출된 예측치(7,565.86건)가 관찰 값에 더 근사하게 나타났다.

2) 시계열분석

빈도 자료가 선형이지 않고 큰 폭으로 변동할 때 사용되는 시계열분석을 사용하여 2014년 7월의 값을 산출하였다. 이때 사용된 모형은 시계열분석법에서 가장 대표적으로 사용되는 ARIMA모형으로 동정 모형은 (0, 0, 1)모델이다. 이 동정 모형은 잔차 검증을 위한 Ljung-Box 검증에서 통계량이 18.17로 df=17에서 유의미수준 p〉.05 를 충족하여 모형이 자료를 설명하기에 가장 적절한 것으로 나타났다. 그림 4는 잔차의 자기상관계수에 대한 코레로그램(correlogram)으로 모든 잔차가 신뢰한계 내에 들어 있으므로 이러한 잔차는 백색잡음(white noise)로 생각된다.

ARIMA 모형의 추정된 모수 값은 표 5와 같다. 그림 또한 이 모형을 통해 예측할 수 있는 그래프는 추정된 모수 값을 통해서 나타난 2014년 7월의 예측치는 8,002.60건 이었다. 95%신뢰 수준에서 상한 값은 12,302.21이며 하한 값은 3,702.99였다. 그림 5는 실제 관찰된 빈도의 그래프와 ARIMA 모형으로 산출된 그래프의 비교이다. 그래프를 살펴봤을 때 관찰 빈도와 예측빈도가 유사하게 보이지만, 2014년 7월 실제 자

료값(7,422건)과 비교해 볼 때, 마코프체인 프로세스를 통해 산출된 예측치(7,565.86)가 관찰 값에 비교적 근접하게 나타났다.

표 5

ARIMA 모형의 추정된 모수 값

모형		추정값	표준오차	t
ARIMA 모형	상수	8220.61	717.48	11.46***
	시차1	-.76	.16	-4.75***

*** $p<.001$

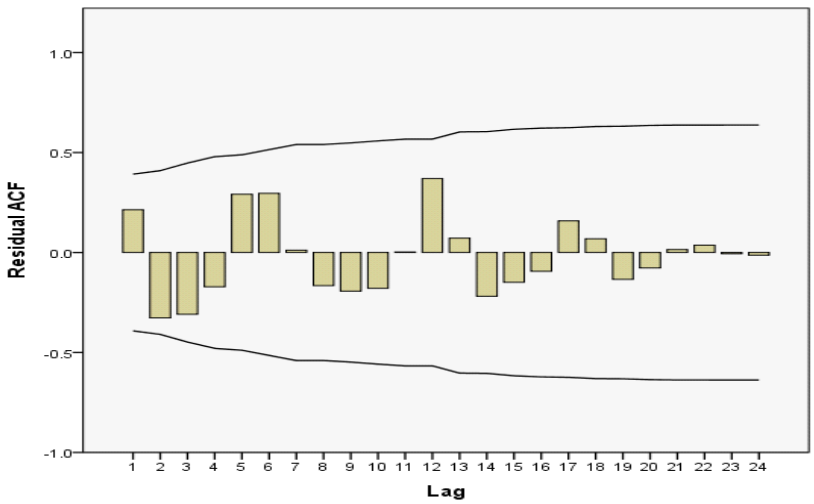


그림 4. 잔차의 자기상관계수에 대한 코레로그램(correlogram) : 히스토그램은 자기상관계수를 나타내며 선은 신뢰한계를 나타냄. 모든 자기상관계수가 신뢰한계 내에 있어 백색잡음(white noise)로 생각된다.

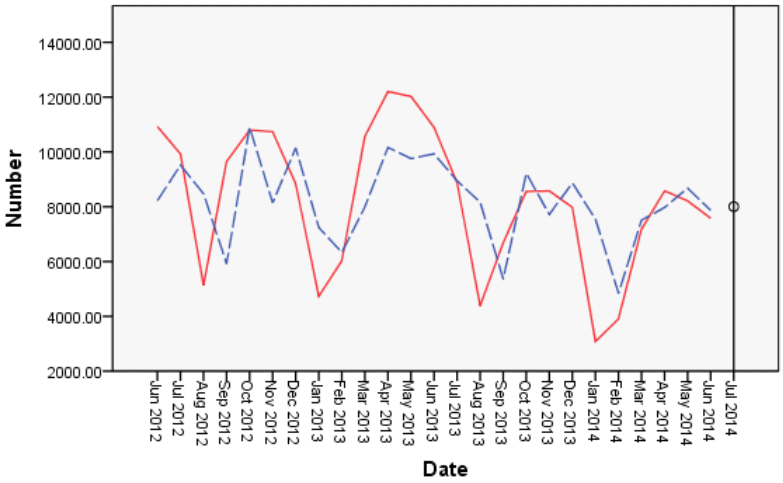


그림 5. 학교폭력 빈도에 대한 시계열 그래프 : 진선은 관찰빈도를 나타내는 선이며, 점선은 ARIMA 모형으로 조정된 선을 나타낸다. 오른쪽 작은 동그라미는 2014년 7월의 예측치(8,002.60건)를 나타낸다.

3) 마코프체인 프로세스의 상태집합의 조정

본 연구에서는 상태집합을 3단계로 구분하여 확률 값을 산출하였는데, 상태집합의 수를 조금 세분화한 것과 비교할 때 확률 값에 어떤 차이가 나타나는지 살펴보고자 하였다. 이를 위해 상태집합을 5개로 구분하였는데, 이를 전체 빈도 분포에서 백분위 20만큼씩 구분하여 집합을 구분하였다. 이는 빈도분포의 양상이 정규분포하여(그림 2, a), 5개 수준의 집단을 구분할 때 동일한 크기의 백분위가 가장 적절하다고 판단하였기 때문이다.

5개의 각 상태에서 다른 상태로 전이되는 횟수를 구하고 이를 바탕으로 상태 전이 행렬식을 식(14)과 같이 구하였다.

$$\begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{array}\begin{array}{c} \begin{array}{ccccc} S_1 & S_2 & S_3 & S_4 & S_5 \end{array} \\ \begin{bmatrix} 1 & 3 & 0 & 1 & 0 \\ 1 & 0 & 2 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 \\ 2 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 3 & 2 \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{array}\begin{array}{c} \begin{array}{ccccc} S_1 & S_2 & S_3 & S_4 & S_5 \end{array} \\ \begin{bmatrix} 0.20 & 0.60 & 0.00 & 0.20 & 0.00 \\ 0.25 & 0.00 & 0.50 & 0.25 & 0.00 \\ 0.17 & 0.33 & 0.33 & 0.17 & 0.00 \\ 0.40 & 0.00 & 0.20 & 0.00 & 0.40 \\ 0.00 & 0.00 & 0.00 & 0.60 & 0.40 \end{bmatrix} \end{array} \quad (14)$$

초기 확률을 산출하기 위해서 이전처럼 최근 5개월간의 학교폭력 데이터를 사용하였다. 산출된 초기 확률 값은 아래 식 (15)와 같다.

$$P = (0.2,\ 0.4,\ \ 0.4,\ \ 0.0,\ 0.0) \quad (15)$$

이 초기확률값과 식(14의) 상태전이 값을 통해 다음 달에 발생될 확률 값은 (0.21, 0.25, 0.33, 0.21, 0.00)으로 나타났다. 즉 S_3 가 나타날 확률이 0.33으로 가장 컸다. 2014년 7월의 빈도를 예측하기 위해 이 확률 값을 대입해 보면 전체 빈도의 평균치(8,211.54)에 대입했을 때, 2,709.05, 최대치(12,203)에 대입했을 때, 4,026.99로 나타났다. 이러한 결과는 상태집합을 3집단으로 구분했을 때 실제 자료를 예측하는데 더 근접한 것으로 보인다.

Ⅳ. 논 의

최근의 학교폭력 문제는 정도의 심각성과 형태의 다양화로 지속적인 관심이 필요한 청소년 문제 중의 하나이다. 이에 정부에서는 학교폭력 종합대책과 전문 기관의 운영으로 적극 대처하고 있다. 학교폭력을 포함하여 가출, 학업중단, 인터넷 중독 등 많은 청소년문제들은 사회적 파장이 매우 크게 때문에 정부와 지방자치단체, 관련 공공기관에서는 이러한 문제를 최소화하기 위한 노력이 많이 필요하다. 따라서 이러한 현상을 적절히 예측하고 관리할 수 있는 예측 체계에 대한 논의가 활발히 진행되고 있다.

하지만 청소년문제 문제 예방을 위한 많은 연구들은 문제의 원인을 탐색하는 변인 중심의 연구가 주류를 이루고 있다. 예를 들어 학교폭력의 원인에 있어서 가정의 불화, 교우 관계, 개인의 내적 기질, 지역 사회 환경 등 여러 변인들을 조사하여 학교폭력의 원인을 탐색하는 연구들이 주류를 이루고 있다. 이러한 연구들은 학교폭력의 원인을 설명할 수 있는 다양한 변인을 제안하고 각 개인이 가지고 있는 여러 변인들을 통하여 개인의 학교폭력 발생 정도를 예측할 수도 있다. 하지만 이러한 변인 중심의 인과모형은 변인이 많으면 많을수록 변인들의 관계가 대단히 복잡해지고 예측할 수 있는 오차가 그 만큼 커지기 때문에 특정 현상을 예측하기에는 대단히 많은 노력과 시간이 필요하다.

따라서 본 연구는 학교폭력 문제 예방을 위해 학교폭력 발생 예측 체계를 위한 보다 효율적인 확률 모형을 제안하고자 했다. 이에 예측 모형으로 여러 분야에서 폭넓게 적용되고 있는 마코프 체인 프로세스 확률 모형을 사용하여 발생 상태와 빈도를 예측하고자 하였다. 사용된 자료는 117학교폭력신고센터의 최근 3년간 자료로서 월별 시계열 자료를 마코프 체인 모형으로 분석하였다.

결과는 마코프 체인 모형으로 산출된 '14년 7월의 학교폭력 발생 확률은 관심상태(S_1) 0.27 주의상태(S_2) 0.62 경계상태(S_3) 0.11로 나타나 주의 상태를 예측하였다. 실제 '14년 7월에 관찰된 학교폭력 발생 빈도는 7,422건으로 주의 상태에 포함되어 있다. 또한 예측 빈도를 산출하기 위해서 산출된 확률 값에 관찰된 빈도의 평균과 최대 값을 적용하였다. 그 결과 산출된 값은 평균의 경우 5,091.15건, 최대값의 경우 7,565.86건으로 예측되어 최대 값으로 산출된 예측 빈도가 관찰된 자료와 근접함을 보였다.

또한 자료의 예측을 위해서 흔히 사용되는 회귀분석과 시계열분석에서 예측하는 자료와 이를 비교하였다. 회귀분석에서 예측된 2014년 7월의 예측치는 6,970,76건 이었으며 시계열분석에서의 예측치는 8,002.60건 이었다. 이는 실제 자료에 있어서 마코프체인 프로세스의 예측치가 다른 분석법과 비교할 때 근사치에 접근하고 있다는 것을 보여주고 있다.

그리고 상태집합을 구분하는 임계치 설정에 따른 비교를 위해서 5개 집합으로 구분한 마코프체인 프로세스의 확률 값을 통한 예측치는 빈도의 최대치에 대입했을 때 4,026.99건으로 나타나 집합을 3개로 구분했을 때에 관찰치에 접근하고 있는 것을 보였다.

따라서 본 연구에서 제안한 마코프 체인 프로세스 확률 모형은 불특정 빈도 패턴을 가지고 있는 학교폭력 발생을 예측할 수 있는 하나의 대안이 될 수 있을 것으로 생각한다.

앞서도 설명했지만 마코프 체인 프로세스는 특정 추세를 가지지 않는 시계열을 가진 자료에도 미래의 상태를 예측할 수 있는 적절한 확률 모형으로서 알려져 있으며 공학, 정보학 등 다양한 분야에 적용되어 왔다. 하지만 청소년 관련 분야에서는 아직까지는 생소한 이론이다. 본 연구의 의의는 이러한 미래 예측을 위한 확률 모형을 청소년 분야에 적용했다는 점이다. 청소년 문제와 관련해서는 이를 사전에 예측하고 대비하는 것이 사회적 불안과 이에 소요될 비용을 최소화하는 방법이다. 따라서 청소년 문제를 예측할 수 있는 다양한 시도가 이루어질 필요가 있다. 본 연구는 학교폭력 문제를 포함한 청소년문제를 예측할 수 있는 하나의 확률 모형을 제시했다는 점에서 의의가 있다. 마코프 체인 프로세스가 학교폭력 등 위기청소년에 대한 발생 빈도를 통해 발생 확률을 적절히 예측할 수 있다면 발생 확률에 따라 전문 인력 배치 및 관련 기관의 대응 계획 등 대응 방안을 사전에 마련할 수 있을 것으로 기대한다. 일례로 교육부 2014년 1차 학교폭력 실태조사 분석 결과에 의하면 학교폭력 피해 공간에 대한 분석자료가 있는데, 교실 안이 39.2%로 가장 높았고, 학교 내 장소가 12.9%로 그 다음 순으로 높았다. 학교 밖 공간에서는 놀이터(5.8%), 학원 주변(3.8%), PC방(1.3%)이 높게 나타났는데 이는 학교폭력 발생 정도가 높아지면 매우 주의해서 관찰해야 할 장소로 생각된다. 만약 학교폭력 발생 정도가 정확하게 예측된다면 학교 안에서는 교사 및 전문상담사, 스쿨폴리스와 같은 대응인력이 보다 주의 깊게 대응할 수 있는 사전 지침이 마련될 수 있을 것으로 기대한다. 또한 학교 밖 주의할 공간에서는 경찰과 같은 인력들이 보다 세심하게 주의 공간을 살펴볼 수 있을 것으로 기대한다. 관련해서 최근 빅데이터를 활용한 GIS(geographic information system) 기술이 발전하면서 사용자가 원하는 관련 자료를 지도 위에서 그래프 등으로 시각적으로 살펴볼 수 있게 되었다. 노찬숙과 김동현(2012)은 이러한 기술을 활용하여 범죄발생율을 마코프 프로세스로 예측하고 범죄에 노출될 수 있는 공원, 가로등 없는 길 등 범죄에 취약 공간에 대한 가중치를 적용하여 특정 장소의 지도위에 범죄 발생 예측 정도를 가중치에 따라 달리 표시하였다. 이는 그 지역의 경찰이 범죄 발생 예측 수준에 따라 위험지역에 대해 보다 주의 깊게 살펴 볼 수 있는 지표가 될 수 있다. 따라서 본 연구에서

제안한 확률모형을 이러한 기술에 적용한다면 학교 내 뿐 아니라 학교 밖 공간에서도 학교폭력 대응 인력을 예측 기간에 따라 적절히 배치하고 주의 깊게 대응할 수 있는 정책이 마련될 수 있을 것으로 기대된다.

이를 위해서는 본 연구에서 제안한 확률 모형을 보다 정교화 할 필요가 있다. 우선 본 연구에서는 학교폭력 발생 빈도를 예측하기 위해서 117학교폭력신고센터의 지난 3년 간 월별 신고건수라는 단순한 자료를 사용하였다. 이는 비교적 단순한 자료를 통해 예측 확률 모형의 적용 가능성을 살펴보고자 한 의도였다. 하지만 보다 정교한 예측 확률 모형을 위해서는 보다 세분화된 자료를 사용할 필요가 있다. 주간 혹은 일별 빈도, 지역별 빈도, 학교폭력의 하위 유형별 빈도 등 보다 세분화된 자료를 이용하여 확률 모형을 적용하면 보다 구체적이고 풍부한 정보를 얻을 수 있을 것으로 기대한다. 또한 117학교폭력신고센터 외의 교육부의 학교폭력 실태 조사 통계 등 보다 다양한 자료가 취합되고 분석된다면 발생 예측의 정교성을 더 높일 수 있을 것으로 기대한다. 그리고 위기청소년의 다양한 문제는 서로 중복되고 상관 정도가 매우 높기 때문에 다른 문제와의 상호 발생 빈도를 분석하면 일정한 패턴을 확률 모형으로 예측할 수 있을 것으로 기대한다. 김영갑 등(2006)의 연구에서는 서로 상관있는 두 가지 웜 바이러스의 발생을 마코프 체인 프로세스로 분석하여 어느 웜 바이러스가 다음 달에 발생할 것인지를 비교적 정확히 예측하였다. 서로 상관이 높은 청소년문제들을 선별하고 이를 마코프 체인 프로세스로 예측한다면 여러 청소년문제들 가운데에서 다음 달, 혹은 차주에 어느 청소년 문제가 예측될 수 있는지를 살펴볼 수 있을 것으로 기대한다. 또한 이러한 세부 적인 자료가 지역별로 모아진다면 지역별 청소년정책 수립에도 도움이 될 것으로 기대한다.

본 연구의 제한점 및 향후 필요한 방향은 다음과 같다.

우선 본 연구에서는 학교폭력 상태를 정의하는데 있어 빈도 분포의 양상을 통해 1표준편차 단위로 구분했다. 또한 다른 임계치와 비교하기 위해 5개 집합의 자료와 정확성을 비교하였지만, 학교폭력 상태를 보다 정확하게 정의하기 위해서는 보다 많은 전문가들의 견해가 모아져서 적절한 임계치를 정할 필요가 있다. 김영갑 등(2006)에 의하면 이러한 임계치 설정이 보다 정교화 될수록 예측의 정확성이 제고될 수 있다고 하였다. 본 연구에서는 마코프 체인 프로세스 확률 모형의 적용 가능성을 살펴보고자 하였다는 점에서 임계치 설정에 대해서 다소 임의적인 방식을 사용하였다는 한계가 있다.

두 번째 한계는 초기 확률 값을 구하는 범위를 본 연구에서는 최근 5개월 구간으로 설정하였는데, 이 구간의 설정 부분도 보다 정교하게 설정될 필요가 있다. 이는 초기 확률 값의 설정에 따라 예측의 정확성도 달라질 수 있기 때문이다. 이를 보다 세분화된 학교폭력 빈도를 얻고 그 빈도에 대한 최근 추세를 보다 상세하게 분석한 뒤 초기 확률 값을 구하기 위한 구간을 설정할 필요가 있다.

마지막으로 본 연구에서는 예측에 영향을 줄 수 있는 다양한 오차를 고려하지 못했다는 점이다. 행동과학의 영역에서는 여러 변인들의 복합적인 상호관련성으로 의외의 결과가 나타날 수 있다. 학교폭력 문제를 포함한 청소년 문제는 사실 대단히 복잡한 변인 간 인과관계에 얽혀 있어서 몇 개의 변인으로만 정확히 현상을 예측하기는 매우 어려운 한계를 가지고 있다. 따라서 앞서 지적한바와 같이 상호관련성 있는 주요 변인들에 대한 연구와 함께 예측할 수 있는 모형에 대한 연구가 필요할 것으로 사료된다.

영국의 수리기상학자 Lewis Fry Richardson(1881~1953)은 1910년 5월 20일 오전 7시 전 세계의 날씨를 상세하게 조사한 뒷, 그로부터 6시간 뒤인 오후 1시의 날씨를 체계적으로 예측해내는 작업에 착수했다(Barabasi, 2010). 굉장히 많은 예산과 인력이 투입되었지만 예측의 정확성은 그야말로 형편없었다. 하지만 100여년이 지난 오늘 Richardson의 연구는 기상학 혹은 일기예보의 기초가 되었으며 그를 통해 현재 우리는 일기 예보를 보면서 다음 날을 준비하는 세상이 되었다. 청소년문제에 관해 많은 학자들이 예측 모형을 만들고자 시도를 하고 있다. 본 연구도 그러한 시도에 하나의 도움이 되기를 기대한다.

참 고 문 헌

교육부 (2014). 보도자료: 2014년 1차 학교폭력 실태조사 분석결과 발표. http://www.moe.go.kr/web/106888/ko/board/view.do?bbsId=339&boardSeq=548802에서 2014년 7월 11일 인출.

김영갑, 백영교, 인호, 백두권 (2006). 마코프 프로세스에 기반한 확률적 피해 파급 모델. **정보과학학회논문지: 시스템 및 이론, 33**(8), 524-535.

노찬숙, 김동현 (2012). 마코프 체인 기반의 범죄 발생 위험도 확률지도 생성 모델. **한국정보기술학회논문지, 10**(10), 89-98.

손재환, 이대형, 이현진, 유춘자, 정진선, 김수현 외 (2012). 학교폭력 가·피해 유형분류를 위한 현장 전문가 개념도 분석. **청소년연구, 21**(2), 317-342.

정영석, 정진영 (2012). 마코프 프로세스를 적용한 범죄 발생 예측 방법에 관한 연구. **한국컴퓨터정보학회, 17**(3), 95-103.

조원진 (2014). 보도자료: 하루평균 267건 학교폭력 신고. http://chowonjin.com/220123516481에서 2014년 8월 24일 인출.

Barabasi, A, L. (2010). **버스트: 인간의 행동 속에 숨겨진 법칙** (강병남, 김명남 역.). 서울: 동아시아.

ABSTRACT

A probabilistic model of school violence based on the markov chain process

Son, Jaehwan*

Recently in Korea, school violence has been becoming more complex and varied in form. Therefore, government and public institutions in Korea have need of an early warning and prediction system that is better able to identify adolescents at risk from such things as school violence. This study was carried out for the purpose of suggesting a probabilistic model that can be used to forecast the likely occurrence rate of school violence in Korea. In the pursuit of this purpose, this study made use of the Markov chain process, which can be applied to a great variety of fields such as engineering, and information science. This study proceeded as follows. Firstly, we investigated the frequency of school violence from 2012 to the 2014 in Korea. Second, we applied the Markov chain process to analyze this data. As a result, using the model as proposed in this study, we have been able to identify a model and system which can predict the occurrence probability and occurrence frequency for school violence in Korea.

Key Words: school violence, markov process

투고일: 2014. 12. 15, 심사일: 2015. 2. 16, 심사완료일: 2015. 3. 3

* Konyang University, kiki5048@hanmail.net