

Analysis of internet addiction in Korean adolescents using sparse partial least-squares regression

Jeongseop Han^a · Soobin Park^b · Donghwan Lee^{b,1}

^aDepartment of mathematics, Korea Military Academy;

^bDepartment of Statistics, Ewha Womans University

(Received January 30, 2018; Revised January 31, 2018; Accepted February 1, 2018)

Abstract

Internet addiction in adolescents is an important social issue. In this study, sparse partial least-squares regression (SPLS) was applied to internet addiction data in Korean adolescent samples. The internet addiction score and various clinical and psychopathological features were collected and analyzed from self-reported questionnaires. We considered three PLS methods and compared the performance in terms of prediction and sparsity. We found that the SPLS method with the hierarchical likelihood penalty was the best; in addition, two aggression features, AQ and BSAS, are important to discriminate and explain latent features of the SPLS model.

Keywords: partial least squares, dimension reduction, sparsity, random effects, hierarchical likelihood

1. 서론

인터넷 중독(internet addiction)은 주로 인터넷 사용의 조절 장애로 정의되며 (Young, 1996), 최근 들어 그 심각성에 대한 인식이 커져 세계적으로 관심 받는 질환(disorder)의 하나로 인식되고 있다 (Laconi 등, 2014). 인터넷 사용 인구는 전세계적으로 25억명 가량에 달하고 있으며 (Mok 등, 2014), 우리나라 또한 스마트폰을 비롯한 여러 매체의 높은 보급율을 바탕으로 인터넷의 접근이 매우 용이해진 이후로, 인터넷의 과사용, 나아가 중독 문제로까지 중요한 사회적 이슈가 되고 있다. 특히 청소년들의 과도한 인터넷 사용은 정상적인 일상 생활을 방해하고, 성인이 되어서도 계속 영향을 미칠 수 있다는 점에서 조기 진단과 치료는 매우 중요하다는 보고가 있다 (Cho 등 2008). 이에 인터넷 중독을 정의하고, 중독 여부를 판단하는 인터넷 중독 진단 척도의 개발이 우선적으로 중요하게 연구되어 왔다. 인터넷 중독을 병적 도박 중독 관점에서 보고 개발된 Young (1996, 1998)의 척도를 시작으로 우리나라에서도 Kang과 Oh (2001)의 한국형 인터넷 중독 척도(K-scale)가 개발 되는 등, 전 세계적으로 23개의 언어로 45개 이상의 인터넷 진단 척도가 개발되어 왔다 (Laconi 등, 2014).

This research was supported by the Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (2014M3C7A1062896).

Jeongseop Han and Soobin Park contributed equally to this work.

¹Corresponding author: Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea. E-mail: donghwan.lee@ewha.ac.kr

인터넷 중독을 행위 중독의 하나로 인식을 하면서, 인터넷 중독 진단에서 중독 위험군과 정상군으로 분류된 집단간의 우울, 불안, 충동 등과 같은 여러 정신 질환간의 상관 관계를 살펴보거나 (Lim 등, 2015), 잠재 집단 분석을 통하여 인터넷 중독군의 유형 분류를 통해 정신 질환 척도간의 차이가 있는지 분석하는 연구가 주로 진행되어 왔다 (Mok 등, 2014). Park 등 (2013)은 경로 분석(path analysis)을 통해 인터넷 중독에 행동 활성화/억제 체계가 미치는 영향을 보면서 충동, 불안 등의 매개효과를 검증한바 있다. 하지만, 이 모형에서는 인터넷 중독과 연관성 및 매개효과를 알아보기 위해 여러 정신질환척도 등 중에 각각 하나의 척도만 매개효과로 넣은 모형을 다루고 있다는 단점이 있다. 다양한 설문문항들과 인터넷 중독 척도 간의 잠재적인 인과성을 탐색하기 위해서는 구조 방정식 모형(structural equation model)이 적합하다. 아직 인터넷 중독과 관련한 구조방정식 연구모형은 찾기 어려우나, 페이스북 또는 스마트폰 중독과 여러 척도들과의 연관성을 구조방정식 모형을 통해 연구된 바 있다 (Dhaha, 2013; Kim 등, 2017).

본 연구에서는 우리나라 청소년들의 자가 설문 자료 (Cho 등, 2014)를 기반으로 인터넷 중독 정도와 우울, 불안, 충동성, 분노 등의 질환 정도 등의 선형 연관성을 확인해보고 인터넷 중독 정도에 대한 예측 모형을 수립하기 위해 부분 최소 제곱(partial least-squares; PLS) 회귀 모형을 이용하여 분석하였다. PLS 회귀 모형은 설명 변수의 개수가 많고 서로 상관성이 높을 때 이들의 선형 결합으로 적은 수의 잠재변수를 만들어 설명변수들의 차원을 요약하고, 만들어진 잠재 변수들을 이용하여 종속 변수를 모형화함으로써 해석력과 예측력을 높일 수 있는 회귀 모형 가운데 하나이다 (ter Braak과 de Jong, 1998).

PLS 모형은 설명변수의 개수가 자료의 개수보다 많은($n < p$) 경우에도, 모형화가 가능하다는 장점이 있지만, 잠재변수를 만들때 모든 변수들을 사용함으로 중요한 변수들을 선택하지는 못함으로 최종 모형의 회귀 계수를 해석하는데 어려움이 남을 수 있다. Chun과 Keleş (2010), Lee 등 (2011)은 PLS 모형에 벌점화 방법(penalization method)을 이용하여 차원요약과 변수 선택이 동시에 가능한 희소 부분 최소 제곱(sparse PLS; SPLS) 방법을 제안한 바 있다. 이들 논문에서는 SPLS 방법이 중요한 설명변수들을 효율적으로 선택하면서도 기존의 PLS 모형보다 예측력이 뛰어난 성능을 보여주고 있다. SPLS에서의 희소성(sparsity)은 로딩벡터에 부여되게 되는데, 벌점 함수의 종류에 따라 그 결과와 성능은 달라지게 된다. Lee 등 (2011)에서는 L1 페널티와 함께 Lee와 Oh (2014)의 다단계가능도 기반의 벌점함수를 비교한바 있다. 본 논문에서도 인터넷 중독 척도를 비롯한 여러 질환 척도들이 상관성이 높은 여러 문항으로 구성되어 있어, PLS와 SPLS 모형을 이용하여 분석을 하고, 예측력 비교를 하였다.

본 논문의 2장에서는 벌점 함수를 이용하여 희소성을 부여하는 방법인 희소 부분 최소 제곱법과 그 추정 방법을 소개하고, 3장에서는 우리나라의 청소년 게임 중독 자료에 제시된 방법을 적용하여 비교분석하고, 4장에서 결과를 정리하였다.

2. 희소 부분 최소 제곱 회귀 모형

2.1. 부분 최소 제곱 회귀 모형

PLS 회귀모형은 차원요약법과 다중 선형 회귀 분석을 결합한 방법으로 경제학, 사회과학 등의 분야에서 널리 사용되고 있는 모형이다 (Hulland, 1999). PLS 모형은 먼저 설명변수 행렬 $X \in \mathcal{R}^{n \times p}$ 와 종속변수 행렬 $Y \in \mathcal{R}^{n \times q}$ 를 잠재 성분 분해(latent component decomposition)를 통해, 직교 요인(orthogonal factor)과 적재값/loading)으로 분해한다. 이러한 분해 과정을 통해 X 와 Y 를 다음과 같이 표현한다

$$X = TQ_X^T + E_X, \quad (2.1)$$

$$Y = TQ_Y^T + E_Y, \quad (2.2)$$

Table 2.1. NIPALS algorithm for partial least-squares (Lee *et al.*, 2011)

초기값	$X_{(1)} = X, Y_{(1)} = Y, v_{(1)}$ 은 Y 의 첫번째 열, $k = 1$
1단계	$r_{(k)} \leftarrow X_{(k)}^T v_{(k)} / (v_{(k)}^T / (v_{(k)})), r_{(k)} \leftarrow r_{(k)} / \sqrt{r_{(k)}^T r_{(k)}}$
2단계	$t_{(k)} \leftarrow X_{(k)} r_{(k)}$
3단계	$b_{(k)} \leftarrow Y_{(k)}^T t_{(k)} / (t_{(k)}^T / (t_{(k)})), b_{(k)} \leftarrow b_{(k)} / \sqrt{b_{(k)}^T b_{(k)}}$
4단계	$v_{(k)} \leftarrow Y_{(k)} b_{(k)}$
5단계	수렴할 때 까지 1-4단계를 반복
6단계	$k < K$ 인 경우, $k \leftarrow k + 1. T_{(k-1)} = (t_{(1)}, \dots, t_{(k-1)}),$ $X_{(k)} = (I_n - T_{(k-1)} T_{(k-1)}^+) X, Y_{(k)} = (I_n - T_{(k-1)} T_{(k-1)}^+) Y$ $T_{(k-1)}^+$: $T_{(k-1)}$ 의 무어-펜로즈 역행렬. 1단계로 돌아가 $k = K$ 까지 반복.

여기서 $Q_X \in \mathcal{R}^{p \times K}, Q_Y \in \mathcal{R}^{q \times K}$ ($K < \min(p, q)$)는 각각 X, Y 에 해당하는 적재 행렬 (loading matrix) 이라고 하며 E_X, E_Y 는 각각 X, Y 에 대한 오차 행렬이다. 잠재 성분 행렬 $T \in \mathcal{R}^{n \times K}$ 는 K 개의 방향 벡터를 열로 가지는 가중치 행렬 (weight matrix) $W \in \mathcal{R}^{p \times K}$ 와 설명변수 행렬 X 로 다음과 같이 표현된다.

$$T = XW. \tag{2.3}$$

식 (2.1)-(2.3)를 이용하여 종속변수와 설명변수 간에 회귀 모형은 다음과 같이 요약된다.

$$Y = X\beta + E_Y,$$

여기서 일반 회귀모형과의 차이점은 회귀계수 $\beta = WQ_Y^T$ 로 표현된다는 점이다. 따라서 PLS는 X 와 Y 가 상대적으로 적은 K 개의 성분으로 차원 요약이 잘 되는 경우에 일반 최소 제곱법 (ordinary least square; OLS) 보다 적은 수의 모수로 모형화가 가능해진다. 원하는 회귀 계수를 얻기 위해서는 가중치 행렬 $W = (w_{(1)}, \dots, w_{(K)})$ 에 들어갈 열벡터 $w_{(k)}$ 를 구해야하는데, $Z = Y^T X, S_X = X^T X/n$ 이라 할 때,

$$w_{(k)}^T w_{(k)} = 1, \quad w_{(k)}^T S_X w_{(h)} = 0, \quad \forall h < k$$

의 제약조건 하에서 $w_{(k)}^T Z^T Z w_{(k)}$ 를 최대로 하는 $w_{(k)}$ 로 주어진다. 이러한 $w_{(k)}$ 를 구할 때, 비선형 반복 부분 최소 제곱법 (non-linear iterative partial least squares; NIPALS)라는 알고리즘이 주로 사용된다. 기존의 NIPALS 알고리즘들은 계산 과정 중 보통 $Y_{(k)}^T X_{(k)}$ 의 계산을 요구한다. 이는 p 와 q 가 큰 경우에 계산상의 효율을 저하시키는 요인으로 작용할 수 있으므로, 본 연구에서는 Lee 등 (2011)에서 제안한 수정된 NIPALS 알고리즘을 사용 하였으며, 구체적인 NIPALS 알고리즘은 Table 2.1과 같다.

Table 2.1에서의 알고리즘 수행 후, 가중치 행렬 W 는 다음과 같이 얻을 수 있다.

$$W = R_{(K)} \left(G_{(K)}^T R_{(K)} \right)^{-1},$$

여기서 $G_{(K)} = X^T T_{(K)} (T_{(K)}^T T_{(K)})^{-1}, R_{(K)} = (r_{(1)}, \dots, r_{(K)})$ 이다.

2.2. 회소 부분 최소 제곱법

PLS 회귀분석은 자료의 개수보다 변수의 수가 더 많아 OLS를 사용할 수 없는 경우나 자료의 다중공선성이 큰 경우에 효과적으로 사용할 수 있는 방법이다. 하지만, 모형을 적합함에 있어 모든 설명 변수를

사용하기 때문에 결과를 해석하거나 중요한 변수를 찾는 데 있어 어려움이 있다. PLS 모형에 벌점화 방법을 적용하여, PLS를 수행하면서도 변수를 선택하는 SPLS 회귀 분석 방법들이 제안되었고, 본 연구에서는 Lee 등 (2011)에서 제안한 다단계 가능도(hierarchical likelihood)를 이용한 SPLS 방법을 사용하였다.

Table 2.1의 NIPALS 알고리즘의 1단계와 3단계인 $r_{(k)} = X_{(k)}^T v_{(k)} / (v_{(k)}^T / (v_{(k)}))$ 와 $b_{(k)} = Y_{(k)}^T t_{(k)} / (t_{(k)}^T / (t_{(k)}))$ 은 e_x 와 e_y 를 오차항으로 가지고 주어진 $v_{(k)}$ 와 $t_{(k)}$ 에 대하여 $r_{(k)}$ 와 $b_{(k)}$ 를 구하는 다변량 회귀 분석 문제로 생각할 수 있다:

$$\begin{aligned} X(k) &= v_{(k)} r_{(k)}^T + e_x, \\ Y(k) &= t_{(k)} b_{(k)}^T + e_y. \end{aligned}$$

PLS 방향 벡터 $r_{(k)}$ 에 회소성을 부여하기 위해 Lee 등 (2011)은 다음의 목적함수를 최소로 하는 $r_{(k)}$ 를 구한다:

$$\text{trace} \left\{ \left(X(k) - v_{(k)} r_{(k)}^T \right)^T \left(X(k) - v_{(k)} r_{(k)}^T \right) \right\} + p_\lambda(|r_j|),$$

여기서 $p_\lambda(|r_j|)$ 는 벌점함수로 0이 아닌 계수들이 많이 포함될 수록 목적함수가 커지게 해주는 함수이다. 따라서, 가능한 계수 추정치들에 0이 많이 부여되고 중요한 계수들만 0이 아닌 값들이 남게 되어 변수 선택이 가능해진다. 본 논문의 3절에서는 Lee 등 (2011)에서 성능이 좋다고 알려진 L_1 벌점함수(Tibshirani, 1996), 다단계 가능도를 이용한 벌점함수(Lee와 Oh, 2014)를 이용하여 그 성능을 비교 분석하였다. 두 벌점함수 $p_\lambda(|r_j|)$ 의 구체적인 수식은 Lee와 Oh (2014), Lee 등 (2011)에 잘 나와있다.

3. 자료 분석

3.1. 자료

본 연구에서는 2013년 4-5월 동안 가톨릭대학교 서울성모병원에서 강원도 춘천시 지역 청소년(중학생) 593명을 대상으로 실시한 인터넷게임 이용 실태 조사 설문 자료(Cho 등, 2014)를 이용하였다. Cho 등 (2014)는 본래 표준화된 자가 진단 인터넷 중독 척도를 개발하기 위해 본 자료를 수집하였다. 해당 설문조사는 서울성모병원의 IRB 승인을 받아 진행되었으며, 인터넷 게임 척도 문항 61문항외에도 인터넷 게임 사용 관련 정보(Internet Usage; IU), 우울(Beck Depression Inventory; BDI), 불안(Beck Anxiety Inventory; BAI), 충동성(Barratt Impulsiveness Scale-11; BIS-11), 주의력 결핍(Conners-Wells Adolescent Self-Report Scale-Short form; CASS-S), 행동 활성화/억제(Behavioral Inhibition and Behavioral Activation Scales; BIS/BAS), 충동적 공격성(Buss-Perry Aggression Questionnaire; AQ), 분노(State-Trait Anger Expression Inventory; STAXI), 공격 신념(Beliefs Supporting Aggression Scale; BSAS)의 10가지 종류의 척도들과 그 척도를 구성하는 세부 문항들을 함께 조사하였다(Table 3.1). 본 연구에서는 인터넷 중독 정도를 측정하는 Korean internet addiction test (K-IAT) 항목의 61개 항목의 점수 합을 종속 변수 Y 로 하고 나머지 9개의 범주에서 결측이 없는 변수 187 문항을 설명 변수로 하여 PLS와 SPLS 분석을 실시하였다. 식 (2.1)-(2.3)을 통해 인터넷 중독에 대한 PLS 모형을 Figure 3.1과 같이 도식화하였으며, 일반 회귀모형과 다르게 설명변수(정신질환척도들)와 종속변수(인터넷중독척도) 사이에 저차원의 잠재성분 T 가 있으며, T 의 차원 K 는 정해지지 않은 값이다.

3.2. 분석 방법

여러 차원 축소 방법론들을 비교하기 위하여 인터넷 게임 중독 자료의 부스트랩 표본을 이용한 모의실

Table 3.1. Description of self-administered questionnaires

설문문항 종류 (척도)	측정 내용	문항 수 (척도)
K-IAT : Korean Internet addiction test	인터넷 중독 정도	61 (4점 척도)
IU : Internet Usage	인터넷 게임사용 관련 정보	12 (2-4점 척도)
BDI : Beck Depression Inventory	우울	21 (4점 척도)
BAI : Beck Anxiety Inventory	불안	21 (4점 척도)
BIS-11 : Barratt Impulsiveness Scale-11	충동성	23 (4점 척도)
CASS-S : Connors-Wells Adolescent Self-Report Scale-Short form	주의력 결핍	27 (4점 척도)
BIS/BAS : Behavioral Inhibition and Behavioral Activation Scales	행동 활성화/억제	20 (4점 척도)
AQ : Buss-Perry Aggression Questionnaire	충동적 공격성	29 (5점 척도)
STAXI : State-Trait Anger Expression Inventory	분노	24 (4점 척도)
BSAS : Beliefs Supporting Aggression Scale	공격 신념	9 (4점 척도)

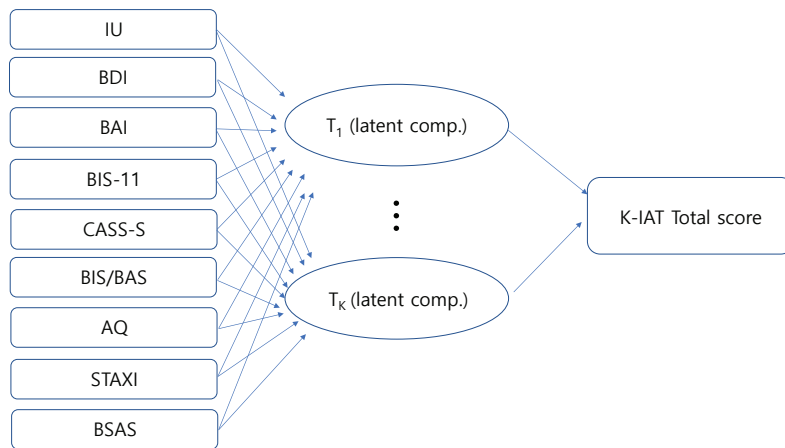


Figure 3.1. Diagram of PLS model. Abbreviations are defined in Table 3.1.

험을 진행하였다. 모의실험은 총 1,000번의 부스트랩 표본을 추출하여 진행하였으며, 각 부스트랩 표본 중, 전체 표본의 80%에 대하여 10-폴드 교차 타당법(10-fold cross validation)을 이용하여 최적 조율 모수들을 찾은 뒤, 희소성 비율로 벌점함수에 의해 0으로 추정된 β 원소들의 비율을 계산하였다. 또한 예측력의 비교를 위해 나머지 20% 표본을 이용하여 제공된 평균 예측오차(root mean-squared prediction error; RMSPE)를 계산하여 예측력을 비교하였다. 본 연구에서 고려된 방법들은 다음과 같다.

- (i) PLS: 부분 최소 제곱 회귀분석
- (ii) SPLS-L1: L_1 벌점 함수를 이용한 희소 부분 최소 제곱 회귀분석
- (iii) SPLS-HL: Lee와 Oh (2014)의 다단계 가능성도 기반 벌점 함수(hierarchical likelihood; HL)를 이용한 희소 부분 최소 제곱 회귀분석

방법 (i)의 경우 `sp1s` R패키지를 이용하였으며, 방법 (ii), (iii)의 경우 Lee 등 (2011)의 `SPLS` 패키지(<http://fafner.meb.ki.se/personal/yudpaw/>)를 이용하였다. 모든 변수는 중심화(centering) 후 분석을 진행하였다.

Table 3.2. Descriptive statistics of PLS and SPLS results

RMSE	PLS	SPLS-L1	SPLS-HL
평균	0.1893	0.1918	0.1881
표준편차	0.0304	0.0303	0.0310
1분위수	0.1676	0.1699	0.1656
중앙값	0.1883	0.1904	0.1864
3분위수	0.2086	0.2110	0.2067

RMSE = root mean-squared error; PLS = partial least-squares; SPLS = sparse PLS; HL = hierarchical likelihood.

Table 3.3. Sparsity of SPLS results

0으로 선택한 변수의 비율			성분의 개수		
PLS	SPLS-L1	SPLS-HL	PLS	SPLS-L1	SPLS-HL
0	0.2193	0.0373	4.58	5.64	4.91

PLS = partial least-squares; SPLS = sparse PLS; HL = hierarchical likelihood.

3.3. 분석 결과

위에서 언급한대로 각 방법론의 비교는 다음과 같은 두 가지 관점에서 비교해 볼 것이다.

- (i) 추정 효율(예측 오차)
- (ii) 희소성 효율(별점 함수에 의해 0으로 추정한 변수의 개수 및 성분의 개수)

3.3.1. 추정 효율 1,000개의 부트스트랩 표본에 대하여, PLS, SPLS-L1, SPLS-HL 세 가지 방법에 대하여 구한 RMSPE는 Table 3.2와 같다. SPLS-L1의 경우 PLS에 비해 예측 오차가 큰데, 변수의 차원을 축소하는 과정에서 전체 자료가 가지고 있던 정보를 효율적으로 요약하지 못하였기 때문인 것으로 판단된다. 반면, SPLS-HL의 경우 SPLS-L1에 비하여 더욱 희소하게 성분을 추정하였음에도 (Table 3.3) 오차가 가장 작은 것으로 나타났다. 즉, SPLS-HL의 경우 세 가지 방법 중 가장 효율적으로 자료의 정보를 요약하여 예측력을 높일 수 있었다. Figure 3.2는 그 결과를 상자 그림으로 나타낸 것이다.

3.3.2. 희소성 효율 이번 절에서는 희소성을 부여한 SPLS 방법에서 어떠한 변수 및 성분들이 선택되었으며, 이들이 선택되는 정도에 따라 살펴볼 것이다. Table 3.3은 1,000번의 부트스트랩 시뮬레이션 희소성 결과의 평균값들을 보여준다.

희소성 제약 조건이 없는 PLS의 경우는 모든 변수들을 사용하므로 희소 비율은 0이고, SPLS-L1은 평균적으로 전체 변수의 22% 정도를 0으로 추정한 반면, SPLS-HL의 경우 평균 4% 정도의 변수만 0으로 추정한 것을 알 수 있었다. 이는 데이터의 특성 상 다양한 출처의 문항들을 넘다보니 잠재성분의 개수가 늘어나 각 성분을 구성하는 계수들은 희소하나 각 성분의 선형결합으로 최종 회귀계수 β 를 추정한 결과이다. 반면 성분의 개수(K) 측면에서는 SPLS-L1가 가장 많은 K 를 고름을 알 수 있다. Figure 3.3(b)를 보면, 시뮬레이션 과정 중 SPLS-L1가 SPLS-HL에 비하여 많은 잠재 성분의 개수를 선택하는 경우가 많음을 알 수 있다. 3.31절에서의 예측 오차 결과에 비추어 볼 때, 0으로 추정한 변수의 정도와 적절한 성분의 선택이 모두 예측력에 영향을 미침을 알 수 있다. 특히 모든 변수들을 모두 사용한 PLS보다 SPLS-HL이 예측력이 좋으면서도 변수 선택을 함으로써 해석의 용이성을 증을 알 수 있다 (Figure 3.3).

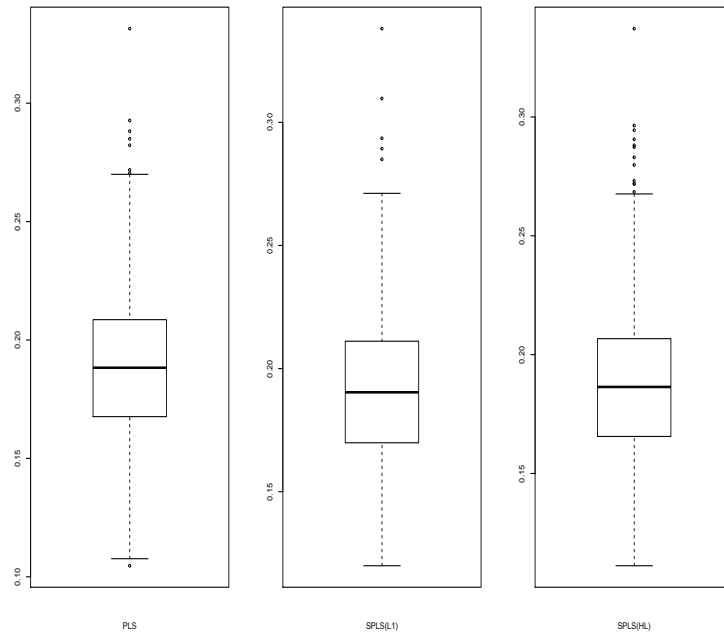


Figure 3.2. Boxplots of RMSE of PLS and SPLS results. PLS = partial least-squares; SPLS = sparse PLS; HL = hierarchical likelihood.

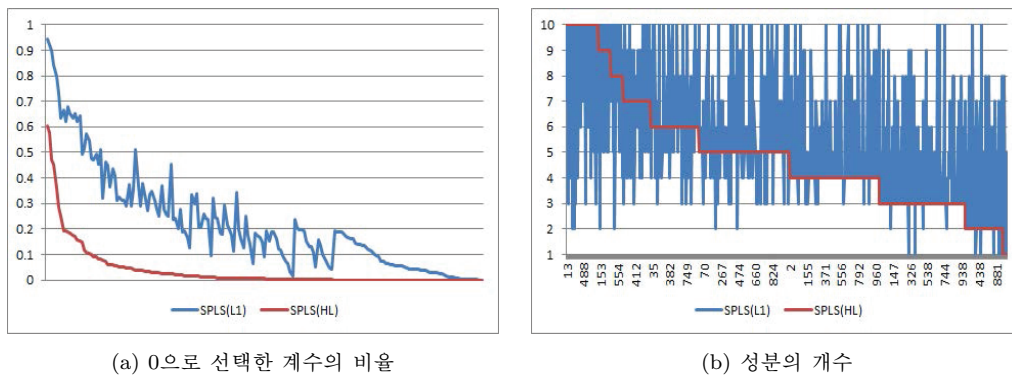


Figure 3.3. Bootstrap simulation. SPLS = sparse partial least-squares; HL = hierarchical likelihood.

3.3.3. 결과 해석 예측 오차면에서 가장 뛰어난 SPLS-HL 방법을 이용하여 K 와 벌점 함수의 조절 모수 λ 를 10-폴드 교차 타당법을 이용하여 최적 조율 모수를 찾고, 다시 전체 자료를 이용하여 적합하여 결과를 보았다. $K = 2$ 가 선택되었으며, SPLS-HL의 희소성 성질에 의해 인터넷 중독 정도를 설명하는데 있어서 최종 모형에서 선택되지 않고 탈락한 21개의 설명변수들을 확인할 수 있다 (Table 3.4). 충동성을 뜻하는 BIS-11 측도는 모형에 모두 선택된 반면에 주의력 결핍과 관련된 CASS-S는 4문항이 제거되었다. 실제로 인터넷 중독과 의미있는 연관성이 없거나, 상관성이 높은 다른 문항과의 다중공선성으로 인해 제거된 것으로 판단된다.

Table 3.4. Omission of irrelevant variables from the SPLS-HL results

변수	문항 번호	설문 문항
IU	11	인터넷 사용 문제 때문에 상담을 받거나 치료를 받은 적이 있나요?
	12	나는 사람들에게 대한 관심이 완전히 없어졌다.
BDI	16	나는 평소보다 몇 시간이나 일찍 깨고 한번 깨면 다시 잠들 수 없다.
	19	나는 현재 음식 조절로 체중을 줄이고 있는 중이다.
BAI	2	나는 흥분된 느낌을 받는다.
	19	나는 가끔씩 기절할 것 같다.
	21	나는 땀을 많이 흘린다
CASS-S	2	나는 내가 따를 수 있는 규율이라면 따르려 한다.
	7	나는 경찰과 자주 접촉하게 된다.
	24	다른 사람의 물건을 부순다.
	26	나는 식사시간 중 가만히 앉아있기가 힘들다.
BIS/BAS	3	나는 친구들에 비해서 별로 두려움이 없는 편이다.
	10	나는 흔히 '아무 것도 나를 방해할 순 없어'라는 식으로 접근한다.
	20	새로운 일이 재미있을 것으로 생각되면, 나는 언제나 그것을 시도한다.
AQ	11	나는 다른 사람과 의견 충돌이 잦다.
STAXI	1	나는 화를 참는다.
	4	나는 사람들에게 인내심을 갖고 대한다.
	19	나는 욕을 한다.
	20	나는 참고 이해하려고 노력한다.
BSAS	5	누구도 살해당할 이유가 없다.
	6	다른 사람들이 그렇게 한다고 해서 누군가에게 상처를 주는 것은 정말 옳지 못하다.

Abbreviations are defined in Table 3.1. SPLS = sparse partial least-squares; HL = hierarchical likelihood.

탈락한 21개의 설명변수들을 제외하고, 166개의 설명변수만으로 식 (2.3)에서 추정된 W 를 이용하여 잠재성분행렬의 추정값 $\hat{T} = X\hat{W} = (X\hat{W}_1, X\hat{W}_2) = (\hat{T}_1, \hat{T}_2)$ 를 추정할 수 있다. Figure 3.4에서 두 잠재성분 \hat{T}_1 과 \hat{T}_2 는 직교하며, 인터넷 중독 점수인 Y 와 선형적인 관계에 있음을 확인할 수 있어 모형 적합이 잘 이뤄졌음을 알 수 있다. 두 잠재성분 모두 Y 에 양의 상관관계를 가지고 있어, Y 를 설명하는데 있어서 T_1 과 T_2 를 구성하는 가중치인 \hat{W}_1 과 \hat{W}_2 값들을 각 설문 문항별로 살펴보면 T_1 과 T_2 의 특징을 구분할 수 있다. Figure 3.5는 \hat{W}_1 과 \hat{W}_2 의 산점도를 나타낸다. (a)는 AQ에 해당하는 문항들만 구별지어 표시하고, (b)는 BSAS에 해당하는 문항들만 구별지어 표시하였다. 대체로 모든 문항들에 대해 첫 번째 잠재성분에 해당하는 W_1 값들은 양수로 추정이 되었지만, W_2 의 경우에는 AQ는 대부분 음수, BSAS는 대부분 양수임을 알 수 있다. AQ는 개인의 공격성향과 관련이 높고, BSAS는 자존감, 이미지 향상 및 공격행동의 정당성 등과 관련이 있다. 두 번째 잠재성분점수인 T_2 는 AQ와 BSAS 값의 형태에 따라 잠재성분 행렬이 달라질 수 있어, 비슷한 수준의 인터넷 중독 정도라고 하더라도 두 척도간의 차이를 고려하였을 때, 그 유형이 달라질 수 있음을 알 수 있다.

4. 결론 및 논의

변수가 수가 자료에 비해 많은 자료의 분석에 모든 변수들을 이용하여 적합한 결과는 해석에 어려움이 있기 때문에 회소 부분 최소 제곱법을 도입하여 변수 및 성분을 축소, 선택하였다. 이를 실제 자료인 우리나라 청소년 게임 중독 자료에 적용하였으며 그 결과 그 중 계층적 가능성도 별점 함수를 도입한 회소 부분 최소 제곱법(SPLS-HL)이 예측력이 가장 좋음을 알 수 있었다. 고려할 수 있는 여러가지 정신질환

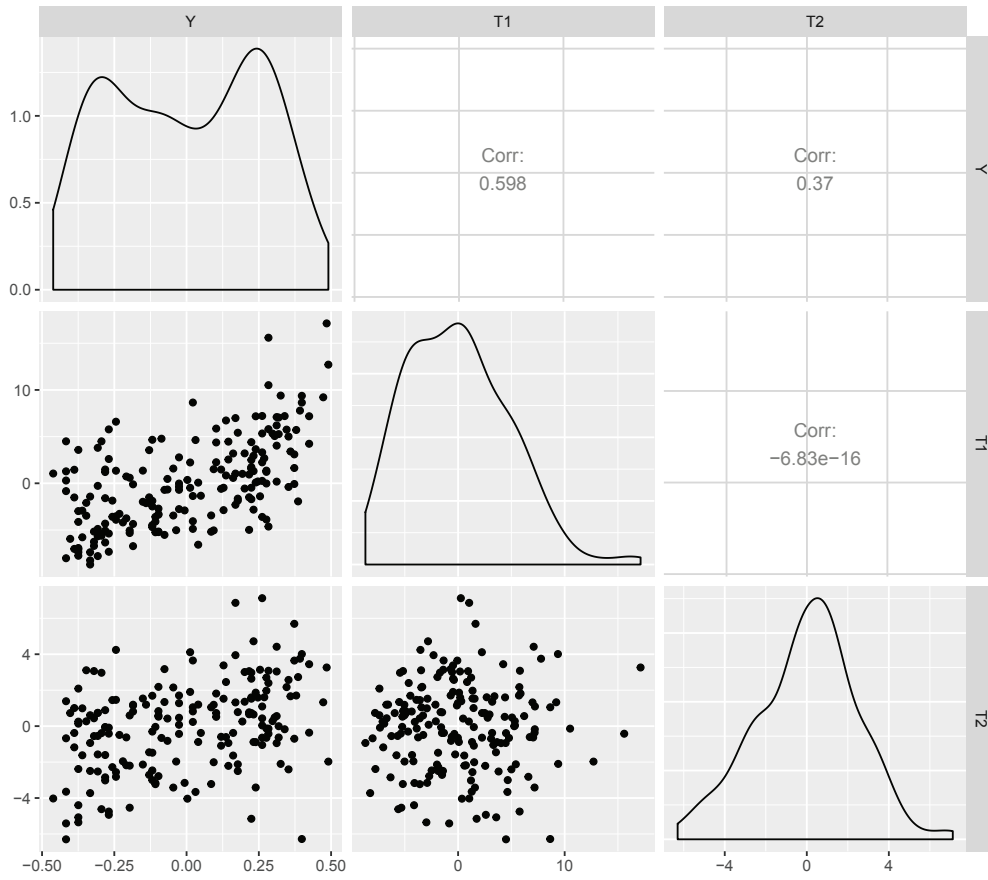


Figure 3.4. Scatter plot matrix among Korean internet addiction test and two latent variables.

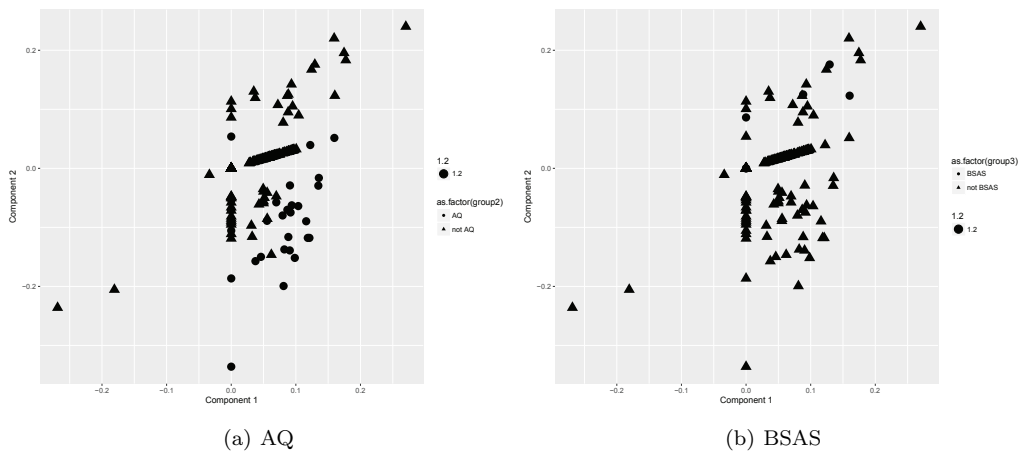


Figure 3.5. Scatter plot of weight vector, W of SPLS-HL. SPLS = sparse partial least-squares; HL = hierarchical likelihood; AQ = Buss-Perry Aggression Questionnaire; BSAS = Beliefs Supporting Aggression Scale.

측도들을 모든 변수들들 활용하여 인터넷 중독 정도를 예측하는 것이 가능함을 보였으며, 예측 뿐만 아니라 적합한 잠재 성분의 계수들의 부호를 통해 표본을 잠재성분별로 유형화가 가능하다는 장점도 얻을 수 있었다. 특히 본 인터넷 중독 자료분석에서는 충동적 공격성(AQ)과 공격 신념(BSAS) 두 설문 점수들에 따라 잠재 성분을 구성하는 계수들의 부호가 달라져 잠재성분 추정치의 정도에 따라 환자군을 유형화 할 수 있다. 이는 인터넷 중독 환자들을 해당 유형별로 맞춤형 예방 또는 치료하는데 반영할 수 있을 것으로 기대된다. PLS 및 SPLS 모형은 잠재성분을 설명변수들의 선형 모형을 표현하고 있다. 또한 모형에서의 잠재성분간의 연관성은 현재 고려하고 있지 않아, 이는 구조방정식 모형을 통해 해결 될 수 있으며, 의미 있는 후속 연구 주제가 될 것이다.

References

- Cho, H., Kwon, M., Choi, J.-H., Lee, S.-K., Choi, J.-S., Choi, S.-W., and Kim, D.-J. (2014). Development of the internet addiction scale based on the internet gaming disorder criteria suggested in DSM-5, *Addictive Behaviors*, **39**, 1361–1366.
- Cho, S. C., Kim, J. W., Kim, B. N., Lee, J. H., and Kim, E. H. (2008). Biogenetic temperament and character profiles and attention deficit hyperactivity disorder symptoms in Korean adolescents with problematic Internet use, *CyberPsychology and Behavior*, **11**, 735–737.
- Chun, H. and Keleş, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection, *Journal of the Royal Statistical Society. Series B*, **72**, 3–25.
- Dhaha, I. S. Y. (2013). Predictors of Facebook addiction among youth: a structural equation modeling, *Journal of Social Sciences*, **2**, 186–195.
- Hulland, J. S. (1999). Use of partial least squares (PLS) in strategic management research: a review of four recent studies, *Strategic Management Journal*, **20**, 195–204.
- Kang, M. C. and Oh, I. S. (2001). Development of Korean internet addiction scales, *Korea Journal of Youth Counseling*, **9**, 114–135.
- Kim, E., Cho, I., and Kim, E. J. (2017). Structural equation model of smartphone addiction based on adult attachment theory: mediating effects of loneliness and depression, *Asian Nursing Research*, **11**, 92–97.
- Laconi, S., Rodgers, R. F., and Chabrol, H. (2014). The measurement of internet addiction: a critical review of existing scales and their psychometric properties, *Computers in Human Behavior*, **41**, 190–202.
- Lee, D., Lee, W., Lee, Y., and Pawitan, Y. (2011). Sparse partial least-squares regression and its applications to high-throughput data analysis, *Chemometrics and Intelligent Laboratory Systems*, **109**, 1–8.
- Lee, Y. and Oh, H. (2014). A new sparse variable selection via random-effect model, *Journal of Multivariate Analysis*, **125**, 89–99.
- Lim, J.-A., Gwak, A. R., Park, S. M., Kwon, J.-G., Lee, J.-Y., Jung, H. Y., Sohn, B. K., Kim, J.-W., Kim, D. J., and Choi, J.-S. (2015). Are Adolescents with internet addiction prone to aggressive behavior? The mediating effect of clinical comorbidities on the predictability of aggression in adolescents with internet addiction, *Cyberpsychology, Behavior and Social Networking*, **18**, 260–267.
- Mok, J.-Y., Choi, S.-W., Kim, D.-J., Choi, J.-S., Lee, J., Ahn, H., Choi, E.-J., and Song, W.-Y. (2014). Latent class analysis on internet and smartphone addiction in college students, *Neuropsychiatric Disease and Treatment*, **10**, 817–828.
- ter Braak, C. J. F. and de Jong, S. (1998). The objective function of partial least squares regression, *Journal of Chemometrics*, **12**, 41–54.
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO, *Journal of the Royal Statistical Society B*, **58**, 267–288.
- Young, K. (1996). Internet addiction: the emergence of a new clinical disorder, *CyberPsychology and Behavior*, **3**, 237–244.
- Young, K. (1998). *Caught in the Net*, John Wiley & Sons, New York.

회소 부분 최소 제곱법을 이용한 우리나라 청소년 인터넷 중독 자료 분석

한정섭^a · 박수빈^b · 이동환^{b,1}

^a육군사관학교 수학과, ^b이화여자대학교 통계학과

(2018년 1월 30일 접수, 2018년 01월 31일 수정, 2018년 2월 1일 채택)

요약

본 연구에서는 우리나라 청소년의 인터넷 중독 자료를 회소 부분 최소제곱법을 적용 하여 분석하였다. 서울 성모병원에서 수집된 자가보고 방식의 인터넷 중독 척도와 다양한 임상 및 정신 병리학적 설문 문항들을 자료로 활용하였다. 표본의 개수보다 설문 문항의 수가 많은 고차원 자료이며, 각 세부문항끼리는 상관관계가 높아 부분 최소제곱법이 좋은 회귀분석 모형이다. 보다 높은 예측 성능과 해석력을 얻기위해 회소성 제약 조건이 가능한 회소 부분 최소 제곱법을 이용하였고, 2가지 다른 벌칙함수를 이용하여 가장 좋은 방법을 선택하였다. 또한, 제안된 최종 모형을 통해 인터넷 중독이 임상 및 정신 병리학적 척도들로 잘 설명됨을 보이고, 공격성과 관련된 다른 설문 문항이 설문 문항이 모형의 잠재성분을 구별하고 설명하는데 역할을 한다는 유의미한 결과를 도출하였다.

주요용어: 부분 최소 제곱법, 차원 축소, 변량 효과, 계층적 가능성.

본 연구는 한국연구재단을 통해 미래창조과학부의 뇌과학원천기술개발사업으로부터 지원받아 수행되었습니다 (2014M3C7A1062896).

한정섭과 박수빈은 공동 제 1저자로 논문에 동등하게 기여했습니다.

¹교신저자: (03760) 서울특별시 서대문구 이화여대길 52, 이화여자대학교 통계학과.

E-mail: donghwan.lee@ewha.ac.kr