

설문 응답에 대한 신뢰도를 반영한 K-척도에 관한 연구

박혜정¹ · 피수영²

¹대구대학교 · ²대구가톨릭대학교 교양교육원

접수 2012년 11월 12일, 수정 2012년 12월 17일, 게재확정 2012년 12월 24일

요약

정보화시대에 인터넷 중독의 심각성은 정보화 사회의 큰 이슈로 부각되고 있다. 인터넷사용이 급증함에 따라 정보화의 역기능도 증가하고 있어 인터넷 중독은 사회적문제로 대두되고 있다. 초고속 무선인터넷 서비스 보급 및 넷북, 스마트 폰 등의 인터넷 접속기기가 더욱 다양화됨에 따라 인터넷 중독 자가진단 검사 척도인 K-척도의 진단기준도 시대변화에 따라 변화가 요구되었으며 2012년 3월에 K-척도의 진단기준이 변경되었다. 본 논문에서는 2012년 변경된 K-척도의 기준으로 경북지역 대학생들의 인터넷 중독 실태와 K-척도 특징들을 살펴보고자 한다. K-척도에서 중독 진단을 위한 조사 방식은 응답자가 직접 자신의 중독증상을 주관적 판단에 의해 응답하는 설문방식이므로 응답자의 고의적인 사실 숨김으로 인해 응답오차가 발생할 수 있다. 본 논문에서는 응답오차를 줄이기 위해 변경된 K-척도에 응답자에게 설문 응답에 대한 신뢰할 수 있는 신뢰도 값을 추가적으로 입력하여 분석의 신뢰도를 높이고자 한다.

주요용어: 가우시안 커널, 군집분석, 서포트벡터 군집분석, 요인분석, 인터넷 중독.

1. 서론

인터넷 중독은 최근에 컴퓨터가 급속도로 보급되면서 역기능의 하나로 등장한 정신적, 심리적 현상이다. 특히 네트워크 환경과 웹 브라우저의 출현으로 인터넷 접근이 쉬워지면서 사회문제로 부각되었다. 정부의 적극적인 투자와 지원을 기반으로 세계 최고 수준의 초고속 인터넷 망 구축 및 저렴한 고성능 컴퓨터 보급으로 인해 현재 국내 가구의 인터넷 보급률은 81.8%, 컴퓨터 보유율은 81.9%이며 전년대비 소폭 증가한 추세이다. 초고속 무선인터넷 서비스 보급 및 넷북, 스마트 폰 등의 인터넷 접속기기가 더욱 다양화됨에 따라 인터넷 환경의 질적인 개선은 계속 될 것으로 보인다. 따라서 인터넷 접속기기가 더욱 다양화됨에 따라 2006년에 개발한 인터넷 자가진단 검사 척도인 K-척도의 진단기준도 시대적 변화 요인들이 반영되어 2011년 변경되었다. 2012년 한국정보화진흥원에서 확정된 성인용 인터넷 자가진단 K-척도를 보면 2006년의 20개 문항에서 15개의 문항으로 변경되었다. 인터넷 중독과 관련된 연구는 청소년과 성인을 대상으로 한 연구는 많지만 대학생을 대상으로 한 연구는 부족한 실정이다 (Choi와 Ha, 2011; Jung 등, 2012; Kim과 Kim, 2012). 본 논문에서는 2012년의 인터넷 중독 자가진단 K-척도에 대한 평가 및 대학생들의 인터넷 중독 실태를 분석하고자 한다.

2012년 인터넷 중독 자가진단 K-척도는 중독 진단을 위한 조사 방식이 응답자가 직접 자신의 중독증상을 주관적 판단에 의해 응답하는 설문방식이므로 응답자의 고의적인 사실 숨김으로 인해 응답오차가

¹ 교신저자: (712-714) 경북 경산시 진량읍 대구대로 201, 대구대학교, 초빙교수.

E-mail: hyjpark@daegu.ac.kr

² (712-702) 경북 경산시 하양읍 금락1리 330, 대구가톨릭대학교 교양교육원, 조교수.

생길 수 있다. 본 논문에서는 설문 응답의 응답오차를 줄이기 위해 2012년 인터넷 중독 자가진단 K-척도 설문에 응답자 자신의 설문 답변에 대한 신뢰할 수 있는 신뢰도 값을 입력할 수 있게 설문 문항을 추가하였다. 본 논문에서는 2012년 인터넷 중독 자가진단 K-척도에 설문 응답자의 신뢰도 값을 추가적으로 입력하여 응답 오차를 줄이며 분석의 신뢰도를 높이고자 한다.

설문은 경북지역 4년제 대학에서 컴퓨터 관련 교양과목을 수강 중인 457명의 대학생을 대상으로 조사하였다. 설문 응답에 대한 신뢰할 수 있는 신뢰도 값이 70%이상인 경우에는 신뢰도 분석 시 신뢰도 값이 높을 뿐만 아니라 분석대상 인원도 적절하게 포함되어 설문 응답에 대한 신뢰도 값이 70% 이상인 417명의 자료를 분석대상으로 이용하였다. 대학생들의 인터넷 중독 자가진단 K-척도가 어떤 특징이 있는지는 판별분석과 요인분석을 통해 간단하게 살펴볼 수 있다. 판별분석과 요인분석은 분석을 하기 위해 각 분석마다 제약 조건들을 만족해야 타당한 분석이라 할 수 있다. 그러나 인터넷 중독 자가진단 K-척도를 통해 측정된 자료들은 정규분포를 하지 않으며 집단별 측정 개수도 동일하지 않다. 또한 판별분석으로 집단들을 분류하며 분석할 때에는 종속변수에 의해 범주화되는 집단들의 분산-공분산행렬이 동일해야 한다는 조건을 만족해야 한다. 본 논문에서 사용되는 자료는 종속변수에 의해 범주화되는 집단들의 분산-공분산행렬이 동일해야 한다는 조건이 만족하지 않는다. 이런 경우에는 비모수적인 방법으로 해법을 찾아야 한다. 이런 제약들을 최대한 반영하기 위해 비모수적인 방법으로 집단을 분류하는 방법을 적용하고자 한다. 비모수적인 분류 방법으로 분류에 로버스트하다고 알려진 서포트벡터기계 (support vector machine; SVM) 기법이 있다. SVM과 관련된 연구는 Vapnik (1995, 1998), Suykens (2000), Suykens와 Vanderwalle (1999, 2001), Hwang (2010, 2011), Shim (2011), Gjorgii 등 (2009)을 통해 알 수 있다. 본 논문에서는 서포트벡터 기법을 반영한 군집분석을 통해 각 집단의 분류특징을 살펴보고자 한다.

본 논문은 다음과 같이 구성하였다. 먼저 2절에서는 관련연구로 변경된 인터넷 중독 자가진단 K-척도 프로그램과 분석방법을 소개하였으며, 현 시점의 인터넷 활용도를 점검하였다. 3절에서는 인터넷 중독 자가진단 K-척도를 통한 대학생들의 인터넷 중독 실태조사 및 K-척도의 특징을 분석하였다. 마지막 4절에서는 본 논문에 대한 결론 및 향후과제에 대해 기술하였다.

2. 관련연구

2.1. 인터넷 중독 자가진단 K-척도와 K-척도의 요인별 분류

2012년 한국정보화진흥원에서 개발한 성인용 인터넷 중독 자가진단 K-척도는 전체 15개 문항으로 구성되어 있으며 15개 문항을 다시 4개의 요인으로 분류할 수 있다. 4개의 요인으로는 일상생활장애 (5문항), 가상세계지향 (2문항), 금단 (4문항), 내성 및 몰입 (4문항) 으로 구성되어 있다. 상세한 문항은 Table 2.1에 정리되어 있다. 설문의 각 문항은 ‘전혀 그렇지 않다 (1)’, ‘때때로 그렇다 (2)’, ‘자주 그렇다 (3)’, ‘항상 그렇다 (4)’로 4점 리커트 척도로 측정하도록 구성되어 있으며, 설문 응답의 총합은 60점이다. 문항 9, 10, 13, 14는 역문항으로 측정한다.

집단별로 분류해 보면 고위험 사용자군은 두 가지 방법으로 분류할 수 있다. 첫 번째는 일상생활장애, 금단, 내성의 세 하위 요인이 모두 표준화점수 (T 점수)가 71점 이상이거나 총점이 표준화점수 71점 이상인 집단이며, 두 번째는 문항 총점 42점 이상 이거나 일상생활장애 14점 이상, 금단 12점 이상, 내성 13점 이상인 경우 고위험 사용자 군으로 분류된다. 잠재적 위험 사용자군은 고위험 사용자 군에 속하지 않으면서 일상생활장애 요인이 표준화 점수 66점 이상이거나 총점이 표준화 점수 66점 이상이며, 설문 문항 관점에서는 총점이 39점 이상이고 41점 이하이거나 일상생활장애가 13점 이상인 경우에 분류된다 (Kim과 Kim, 2011).

Table 2.1 Factor analysis of K-scale

Factor	No	Items
Factor1 (Disturbance of daily living)	1	The ability to work or school grades dropped due to excessive use of Internet.
	5	I'm tired due to the use of the Internet, so I sleep in classes(work).
	6	I didn't do what needs to be done while I use internet.
	10	I do not internet enough tired.
	13	Even if use the internet,I am able to perform properly about plan things.
Factor2 (Virtual world orientation)	2	I have more confidence while using the Internet.
	7	I feel better and easily excited when use the Internet.
	3	If I do not use the Internet, I wonder what else was going on. so I can not do other things.
Factor3 (Withdrawal)	8	If it does not run correctly when use the internet, I get annoyed.
	11	If I do not have access to the Internet, I feel impatient.
	14	I am not anxious without use the Internet.
	4	I think, I should stop using a Internet, while I am using it, but I continue use it.
Factor4 (Tolerance and preoccupation)	9	I can adjust the time using the internet.
	12	If I start Internet, I use the a lot more time than I thought at first.
	15	I constantly think to cut down time spent on a Internet.

2012년 인터넷 중독 자가진단 K-척도는 요인별 분류를 위해 사전에 준비된 예비 문항에서 신뢰도 분석을 통해 선정된 문항들로 구성하였으며, 인터넷 중독 자가진단 K-척도의 요인구조를 파악하기 위해 AMOS 7.0을 사용하여 Figure 2.1의 모형으로 확인적 요인분석을 실시하여 요인들을 분류하였다. 인터넷 중독 자가진단 K-척도의 확인적 요인분석을 위해 잠재변수로 설정된 각 하위요인별 관측변수들은 하위요인에 포함되는 문항들을 2개 혹은 3개로 묶어서 설정되어 있다.

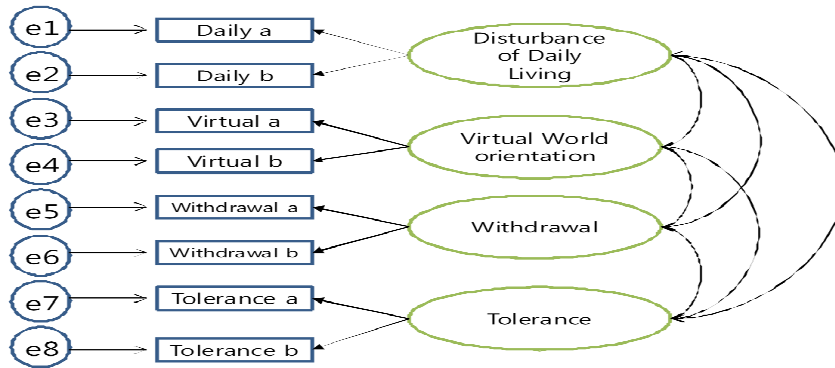


Figure 2.1 Confirmatory factor analysis model of K-scale

먼저 확인적 요인분석 모형을 살펴보면 일상생활장애에서는 1, 5, 10번 문항을 하나의 관측변수로, 6번 13번 문항을 묶어서 관측변수로 설정되어 있으며, 가상세계지향성은 2, 7번 각각을 서로 다른 관측변수로 설정되어 있다. 금단은 3, 8번을 하나로, 11, 14번을 하나로 묶어 각각의 관측변수로 설정되어 있으며, 내성은 4, 9번을 하나로 묶고, 12, 15번을 하나로 묶어 서로 다른 관측변수로 설정되어 있다. 2012년 한국정보화진흥원의 인터넷 중독 자가진단 K-척도의 확인적 요인분석 모형의 적합도는 적합도 지수 NFI, TLI, CFI, RMSEA (.959,.935,.967,.083)로 측정되었다. 측정된 적합도 지수들이 판정 기준치를 충족하여 2012년 인터넷 중독 자가진단 K-척도는 일상생활장애, 가상세계지향성, 금단, 내성의 4 요인 및 그에 해당하는 하위 문항들로 구축되었다. 2012년 인터넷 중독 자가진단 K-척도가 어떤 특징이 있는지는 판별분석과 요인분석을 통해 간단하게 살펴볼 수 있다. 그러나 판별분석과 요인분석은 분석을 하기 위해 각 분석마다 제약 조건들을 만족해야 타당한 분석이라 할 수 있다. 그러나 인터넷 중독 자가진단 K-척도를 통해 측정된 자료들은 정규분포를 하지 않으면 집단별 측정 개수도 동일하지 않다. 이런

제약들을 최대한 반영하기 위해 비모수적인 방법으로 집단을 분류하는 방법을 적용하고자 한다. 비모수적인 분류 방법으로 분류에 로버스트하다고 알려진 SVM 기법이 있다. 본 논문에서는 서포트벡터 기법을 반영한 군집분석을 통해 각 집단의 분류특징을 살펴보고자 한다.

2.2. 서포트벡터 기법의 군집분석

통계적 학습이론 (statistical learning theory)에 기반을 둔 SVM은 주어진 문제를 항상 전역적 최적의 해를 구하기 때문에 패턴인식 분야에서 활용도가 높다. 서포트벡터 클러스터링 (support vector clustering; SVC) 알고리즘의 자료들은 커널함수를 이용해 데이터 공간에서 고차원 특징 공간 (feature space)으로 사상 (mapping)된다. SVC는 특징 공간에서 데이터들의 상 (image)을 둘러싸고 있는 가장 작은 구 (sphere)를 찾으며, 이 구는 데이터 공간으로 사상된다. SVC는 데이터 점들을 둘러싸고 있는 등고선 (contour)의 집합을 형성하게 된다. 이 등고선들은 군집의 경계 (boundary)들로 해석되며, 각 분리된 등고선에 둘러싸여져 있는 점들은 같은 군집으로 할당된다. SVC는 Schölkopf 등 (2001)과 Tax와 Duin (1999)에 의해 공식화 되었다.

d-차원 입력 공간에 존재하는 개체 $x \subseteq R^d$ 로 구성되는 n 개의 데이터 공간의 자료 집합을 $\{\mathbf{x}_i\} \subseteq x$ 라고 하자. 입력자료 \mathbf{x} 는 비선형 변환 (nonlinear transformation) 함수 ϕ 를 사용해서 고차원 특징 공간으로 사상되며 가장 작게 둘러싸고 있는 구를 찾는다.

$$\|\phi(\mathbf{x}_i) - \boldsymbol{\alpha}\|^2 \leq R^2, \quad i = 1, \dots, n,$$

단, $\|\cdot\|$ 는 Euclidean norm, $\boldsymbol{\alpha}$ 는 구의 중심.

ξ_k 는 k -번째 입력 데이터 \mathbf{x}_k 가 구에서 벗어나는 정도를 나타낸다. 라그랑주 (Lagrangian) 함수를 사용하여 식을 정리하면 다음과 같다.

$$L = R^2 - \sum_k (R^2 + \xi_k - \|\phi(\mathbf{x}_k) - \boldsymbol{\alpha}\|^2) \beta_k - \sum \xi_k \mu_k + C \sum \xi_k,$$

여기에서 $\beta_k (\geq 0)$ 와 $\mu_k (\geq 0)$ 은 라그랑주 승수 (Lagrangian multiplier), C 는 벌칙 상수이다. 최적화 문제는 구의 반경 R^2 과 입력 자료가 구를 벗어난 경우에 부과되는 총 벌점 관련항 $\sum \xi_k$ 의 가중치 합 (weighted sum)으로 구성된다. 벌칙 상수 C 는 R^2 과 $\sum \xi_k$ 의 상대적 중요성을 조정하는 상수 (trade-off constant)이다. 최적 해는 라그랑주 함수가 정의되는 확장된 공간에서 안장점 (saddle point)이 되도록 다음 조건을 만족해야 한다.

$$\sum_k \beta_k = 1, \quad \beta_k = C - \mu_k, \quad \boldsymbol{\alpha} = \sum_k \beta_k \phi(\mathbf{x}_k). \quad (2.1)$$

KKT (Karush-Kuhn-Tucker) 조건 (Kuhn과 Tucker, 1951)에 따라 다음의 결과를 얻게 된다.

$$\xi_k \mu_k = 0, \quad (R^2 + \xi_k - \|\phi(\mathbf{x}_k) - \boldsymbol{\alpha}\|^2) \beta_k = 0. \quad (2.2)$$

$\xi_k > 0$ 이고, $\beta_k > 0$ 인 점 \mathbf{x}_k 는 식 (2.2)에 따라 특징 공간 구의 바깥쪽에 놓이며, $\xi_j \mu_j = 0$ 은 한 점이 $\mu_k = 0$ 을 가진다고 할 수 있다. 식 (2.1)에 의해 $\beta_k = C$ 가 되는 자료를 경계 서포트벡터 (bounded support vector; BSV)라고 하며, $\xi_k = 0$ 인 \mathbf{x}_k 는 특징공간의 구 표면이나 내부에 사상된다. 만약 한 점이 $0 < \beta_k < C$ 라면 특징공간의 구 표면에 놓여 있게 됨을 내포하며, 이를 서포트벡터 (support vector; SV)라고 한다. SV들은 군집의 경계 상에 놓이며, BSV들은 군집의 경계 밖에 놓인다. 그리고 다른 모

든 점들은 군집의 경계 안에 존재한다. $C \geq 1$ 일 때는 식 (2.1)에 의해 BSV는 존재하지 않게 된다. 라그랑주 함수를 dual form으로 변환하여 β_j 로 표현하면 다음과 같다.

$$W = \sum_k \phi(\mathbf{x}_k)^2 \beta_k - \sum_{i,k} \beta_i \beta_k \phi(\mathbf{x}_i) \phi(\mathbf{x}_k), \text{ 단, } 0 \leq \beta_k \leq C, k = 1, \dots, n.$$

W 는 Mercer 커널 (Mercer, 1909) $K(\mathbf{x}_i, \mathbf{x}_k) = \phi(\mathbf{x}_i) \phi(\mathbf{x}_k)$ 를 사용하여 다음과 같이 나타낼 수 있다.

$$W = \sum_k K(\mathbf{x}_k, \mathbf{x}_k) \beta_k - \sum_{i,k} \beta_i \beta_k K(\mathbf{x}_i, \mathbf{x}_k). \quad (2.3)$$

따라서 점 \mathbf{x} 에 대해 그 점이 특징공간에 사상된 상과 구 중심과의 거리를 구할 수 있다. 본 논문에서는 군집의 등고선을 산출하기 위해 가우시안 커널 (Gaussian kernel)을 사용하였으며, 식은 다음과 같다.

$$K(\mathbf{x}_i, \mathbf{x}_k) = e^{-q \|\mathbf{x}_i - \mathbf{x}_k\|^2}.$$

여기서 q 는 너비 파라미터이며, \mathbf{X} 의 특징공간에 사상된 상과 구 중심과의 거리 R^2 은 다음과 같다.

$$R^2(\mathbf{x}) = \|\phi(\mathbf{x}) - \boldsymbol{\alpha}\|^2.$$

식 (2.1)을 위의 식에 적용하면 다음과 같다.

$$R^2(\mathbf{x}) = K(\mathbf{x}_i, \mathbf{x}_k) - 2 \sum_k \beta_k K(\mathbf{x}_k, \mathbf{x}_k) + \sum_{i,k} \beta_i \beta_k K(\mathbf{x}_i, \mathbf{x}_k). \quad (2.4)$$

구의 반경 R 은 SV로 구성되어 있으며, 등고선들은 군집의 경계를 형성한다. 즉 SV는 군집 경계 위에 놓여 있으며, BSV는 경계 바깥쪽에, 나머지 다른 자료들은 군집의 경계 안쪽에 놓인다. SVC 기법으로는 Yang 등 (2002)이 제안한 SVC 기반의 근접성 그래프 (proximity graph) 방식인 DD (delanay diagram), MST (minimum spanning tree), k-NN (k-nearest neighbors)와 Lee와 Lee (2005)가 제안한 SVC 기반의 동적 시스템 (dynamical system) 방식인 SEP-CG (stable equilibrium point-complete graph)가 있다. 본 논문에서는 SEP-CG방식을 이용해서 비교 분석한다.

3. 실험 및 결과분석

3.1. 성인 인터넷 중독 자가진단

2012년 한국정보화진흥원에서 개발한 성인용 인터넷 중독 자가진단 프로그램을 경북지역 4년제 대학에서 컴퓨터 관련 교양과목을 수강 중인 학생을 대상으로 설문조사하였다. 전체 설문은 457명의 학생이 응답하였다. 설문에 응답한 457명 중에 설문분석의 신뢰도를 높이기 위해 설문내용 중에 설문 응답에 대한 본인의 신뢰도를 입력하도록 하였다. 설문 응답에 대한 신뢰할 수 있는 신뢰도 값이 70% 이상인 경우에는 신뢰도 분석 시 신뢰도 값이 높을 뿐만 아니라 분석대상 인원도 적절하게 포함되어 설문 응답에 대한 신뢰도 값이 70% 이상인 417명의 자료를 분석대상으로 이용하였다. 설문 응답자들의 기초자료를 보면 학년별로 1학년 225명, 2학년 106명, 3학년 46명, 4학년 40명이며, 성별로는 남자 222명, 여자 192명, 무응답 3명이었다. 거주지별로는 집 234명, 자취 57명, 기숙사 117명, 기타 9명이었다. 설문 응답자들이 수강하는 컴퓨터 교양과목은 수강 학생이 1학년 학생들이 많아 상대적으로 전체 학년 중에 응답자들이 많았다. 먼저 2012년 한국정보화진흥원에서 개발한 성인용 인터넷 중독 자가진단 K-척도

(K-척도)에서 채택하고 있는 문항들에 대한 신뢰도를 살펴보면 다음과 같다. K-척도의 전체 문항에 대한 신뢰도는 0.877이며, 각 요인별 신뢰도는 일상생활장애 (1요인)의 신뢰도는 0.733이었으며, 가상세계지향성 (2요인)은 0.678, 금단 (3요인)은 0.630, 내성 및 몰입 (4요인)은 0.708이었다. 각 요인별로 문항들이 신뢰도가 높은 것으로 나왔다. K-척도에 따라 설문응답자들을 분류해보면 전체 417명 중에 일반사용자군은 360명 (86.3%), 잠재적사용자군은 51명 (12.2%), 고위험사용자군은 6명 (1.4%)으로 분류되었다. K-척도의 결과를 학년별, 성별로 분류해보면 Table 3.1과 Table 3.2와 같다. Table 3.1에서 저학년일수록 잠재적 사용자 군과 고위험 사용자 군이 많음을 알 수 있다. 이는 고학년일수록 취업과 관련된 생활환경의 변화가 인터넷 중독에 조금은 영향력이 있는 것을 감지할 수 있었다. 성별에 대해서는 K-척도 분류 집단 간에 약간의 차이가 있었다.

Table 3.1 The diagnostic results of K-scale for each grade

	Grade1	Grade2	Grade3	Grade4	Total
General user group	190	95	43	32	360
Potential-risk user group	32	9	3	7	51
High-risk user group	3	2	0	1	6
Total	225	106	46	40	417

Table 3.2 The diagnostic result of K-scale for each gender

	Male	Female	Total
General user group	189	169	358
Potential-risk user group	29	21	50
High-risk user group	4	2	6
Total (Non response : 3)	222	192	414

3.2. K-척도에 대한 요인분석과 판별분석

다음으로 요인 분석 결과를 살펴보면 Table 3.3에서 Table 3.5와 같다. 설문 자료를 통해 추출된 요인의 수는 Table 3.3과 같다. Table 3.3에서 고유값 (eigenvalue)이 1보다 큰 요인으로 3개가 있음을 알 수 있으며, 추출된 각 요인의 설명력을 확인할 수 있다. Table 3.4는 공통성(communalities)에 관한 내용으로 각 변수의 초기값 (initial)과 주성분 분석에 의한 각 변수의 추출값 (extraction)이 제시되어 있다. 각 변수의 공통성은 추출된 요인에 의해 설명되는 비율을 나타내며, 일반적으로 공통성이 0.4 이하이면 낮다고 판정한다. Table 3.4에서 문항 5, 15가 낮은 공통성을 보이고 있음을 알 수 있다. Table 3.5를 보면 요인 1에는 문항 1, 2, 3, 7, 11, 요인 2에는 4, 6, 8, 12, 요인 3에는 9, 10, 13, 14와 같이 묶이는 것을 알 수 있다. Table 3.5를 통해 설문 문항들이 연관성이 있는 문항이 많은 관계로 K-척도 기준과는 사뭇 다르게 분류됨을 알 수 있다. 또한 문항 5, 15번은 어떤 요인에도 속하지 않음을 알 수 있다. 요인분석에서 한 가지 특징적인 경향은 설문 문항 중 역문항으로 설계한 문항들이 한 요인으로 묶이는 것을 알 수 있다.

Table 3.3 Total variance explained

Comp.	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative%	Total	% of Variance	Cumulative%	Total	% of Variance	Cumulative%
1	5.640	37.601	37.601	5.640	37.601	37.601	3.147	20.980	20.980
2	1.394	9.293	46.894	1.394	9.293	46.894	2.719	18.129	39.109
3	1.255	8.365	55.259	1.255	8.365	55.259	2.423	16.150	55.259

Table 3.4 Communalities

Items	Initial	Extraction
1	1.000	.480
2	1.000	.619
3	1.000	.484
4	1.000	.674
5	1.000	.403
6	1.000	.646
7	1.000	.536
8	1.000	.418
9	1.000	.679
10	1.000	.567
11	1.000	.536
12	1.000	.616
13	1.000	.643
14	1.000	.636
15	1.000	.351

Table 3.5 Factor analysis results

Items	Factor1	Factor2	Factor3
1	.547	.353	.239
2	.771	.142	.073
3	.621	.255	.181
4	.261	.715	.308
5	.447	.299	.338
6	.282	.733	.170
7	.677	.274	.047
8	.367	.530	-.056
9	.049	.364	.737
10	.209	.043	.722
11	.693	.044	.233
12	.102	.762	.159
13	.055	.396	.695
14	.434	-.103	.661
15	.448	.361	.143

다음으로 대학생들의 인터넷 중독 분류 집단에 대한 요인별 특징을 살펴보기 위해 판별분석을 하였다. 판별분석 시 다중 회귀식을 추정하기 위해 동시입력방식과 단계입력방식을 이용하여 분류결과를 비교하였다. K-척도 집단군에 대한 각 요인별 특징은 Table 3.6에 나타나 있다.

Table 3.6 Technical statistics for each factor of K-scale

Discriminant analysis	Disturbance of daily Living		Virtual world orientation		Withdrawal		Tolerance and preoccupation	
	M	SD	M	SD	M	SD	M	SD
General	9.14	1.999	3.12	.987	6.75	1.789	8.48	2.079
Potential-risk user	13.29	.855	3.86	1.059	8.45	1.983	11.25	1.647
High-risk user	14.33	1.862	4.83	.753	10.33	1.506	14.33	1.366
Total	9.72	2.397	3.24	1.039	7.01	1.933	8.90	2.310

동시입력방식을 먼저 살펴보면, K-척도의 요인별 집단간의 차이검증 결과는 Table 3.7에 정리되어 있다. Table 3.7을 보면 각 요인별로 집단 간의 유의한 차이가 있음을 알 수 있다. 또한 Wilks' 람다는 집단 내 분산/(집단 내 분산+집단간 분산)의 비율로서 집단 간 분산이 집단 내 분산에 비해 클수록 0에 가까워지며 그 반대의 경우 1에 가깝게 나타난다. K-척도의 요인 중 일상생활장애의 값과 내성 및 몰입의 값이 가장 작은 Wilks' 람다의 값을 나타내며, 이는 이 변수의 판별력이 대체로 가장 높을 수 있음을 나타낸다.

Table 3.7 F-test of between groups

Group average test	Wilks' lambda	F	Sig.
Disturbance of daily living	.623	125.448	.000
Virtual world orientation	.910	20.375	.000
Withdrawal	.874	29.965	.000
Tolerance and preoccupation	.764	63.834	.000

Fisher 선형판별함수는 성인용 인터넷 중독 자가진단 프로그램에 참여한 새로운 분류대상에 대해 그 분류대상의 요인변수들의 값을 분류함수에 삽입하여 계산한 결과 큰 값으로 나타나는 집단에 분류시킨다. 동시입력방식을 이용해서 분류된 판별적중률 (hit ratio)은 83.2%로, 원래의 집단 케이스 중 83.2%로 올바르게 분류된 것을 알 수 있다.

Table 3.8 Linear discriminant function between groups

Discriminant function	General user group	Potential-riskuser group	High-riskuser group
Disturbance of daily living	1.578	2.813	2.273
Virtual world orientation	.651	.426	.985
Withdrawal	.616	.521	.760
Tolerance and preoccupation	.759	.789	1.656
(Constant)	-14.621	-27.266	-35.563

Table 3.9에서의 단계입력방식을 살펴보면 전체 요인 중 일상생활장애요인이 제일먼저 채택이 되었으며, 다음으로 내성 및 몰입이 채택되었다. Fisher 선형관별함수는 Table 3.10에 정리되어 있으며, 단계입력방식을 이용해서 분류된 최종 요약표는 Table 3.11과 같다.

Table 3.9 Step-by-step factor selection

Step	Variables entered	Wilks' lambda		
		Statistic	Exact F	Sig.
1	Disturbance of daily living	.623	125.448	.000
2	Tolerance and reoccupation	.604	59.179	.000

Table 3.10 Fisher's linear discriminant function

Classification function	General user group	Potential-risk user group	High-riskuser group
Disturbance of Daily Living	1.906	3.068	2.711
Tolerance and preoccupation	.946	.938	1.899
(Constant)	-13.816	-26.770	-34.140

Table 3.11 Summary table about step-by-step input

		Prediction group			Total
		1	2	3	
Frequency	1	316	42	2	360
	2	0	43	8	51
	3	0	0	6	6
	Total	316	85	16	417
%	1	75.8%	10.1%	0.5%	86.3%
	2	0.0%	10.3%	1.9%	12.2%
	3	0.0%	0.0%	1.4%	1.4%
	Total	75.8%	20.4%	3.8%	100.0%

Table 3.11에서 관별적중률은 87.8%로, 원래의 집단 케이스 중 87.8%로 올바르게 분류된 것을 알 수 있다. 결과적으로 동시입력방식의 결과보다는 단계입력방식의 결과가 더 좋음을 알 수 있다. K-척도를 위한 전체 4개 요인을 사용하는 것보다는 영향력 있는 2개 요인을 이용하여 집단들을 분류하는 것이 더 효과적임을 알 수 있다. 그러나 분석한 관별분석 결과에는 문제점이 있다. 관별분석에 사용되고 있는 자료는 종속변수에 의해 범주화되는 집단들의 분산-공분산행렬이 동일해야 한다는 조건을 충족시키지 못한다. 이런 경우에는 비모수적인 방법으로 해법을 찾아야 한다. 그러나 관별분석에서 중요한 요인들을 찾을 수 있어서 이 모수에 대한 정보 획득이라는 면에서 득이 되는 분석이라 할 수 있다.

3.3. 서포트벡터 기법의 군집분석

다음으로 SVC 기법을 이용하여 인터넷 중독 집단군을 분류하여 보았다. 본 논문에서는 SVC 기법 중 SEP-CG 방법을 이용하였다. SVC를 위한 입력 변수는 판별분석에서 선택한 2개의 요인 (일상생활 장애, 내성 및 몰입)으로 하였다. SVC에서 사용된 커널 모수 σ 는 1.1, 벌칙 상수 C 는 1로 되었다. σ 와 C 는 CV (cross validation) 방식에 의해서 결정되었으며, SV의 개수는 198개였다. SVC의 결과와 기존 SPSS 계층적 군집분석 결과를 판별적중률로 비교해보면 SVC는 84.89%, SPSS 계층적 군집분석은 60.19%로 SVC가 성능이 더 좋을 수 있다. 현 설문자료와 같이 어떠한 가정을 할 수 없을 경우에는 SVC와 같이 비모수적인 방법으로 분석하는 것이 더 정확하다.

Table 3.12 Results of SVC

	Prediction group			Total	
	1	2	3		
Frequency	1	303	46	11	360
	2	1	49	1	51
	3	0	4	2	6
	Total	304	99	14	417
%	1	72.7%	11.0%	2.6%	86.3%
	2	0.2%	11.8%	0.2%	12.2%
	3	0.0%	1.0%	0.5%	1.4%
	Total	72.9%	23.7%	3.4%	100.0%

4. 결론

본 논문에서는 대학생의 인터넷 중독 실태 및 특징을 분석하기 위하여 2012년 성인용 인터넷 중독 자가진단 K-척도에 대한 설문지를 이용하여 경북지역 컴퓨터 교양 과목을 수강하고 있는 대학생을 대상으로 조사하였다. 또한 2012년 인터넷 중독 자가진단 K-척도 설문에 응답자 자신의 설문 답변에 대한 신뢰할 수 있는 신뢰도 값을 입력할 수 있게 설문 문항을 추가적으로 입력하여 응답 오차를 줄이며 분석의 신뢰도를 높이고자 설계하였다. 분석결과 저학년일수록 잠재적 사용자 군과 고위험 사용자 군이 많음을 알 수 있었다. 이는 고학년일수록 취업과 관련된 생활환경의 변화가 인터넷 중독 증상에서 벗어날 수 있게 조금은 영향력이 있는 듯하다. 성별로는 K-척도 분류 집단 간에 약간의 차이가 나타났다. 요인 분석에서 추출된 요인의 수는 3개로 K-척도에 나와 있는 4요인과는 조금 다르게 분류됨을 알 수 있었다. 이는 실험을 통해 설문 문항들 간에 연관성이 있는 문항이 많은 관계로 K-척도 기준과는 사뭇 다르게 분류됨을 알 수 있었다. 또한 문항 5, 15번은 어떤 요인에도 속하지 않음을 알 수 있었다. 문항 5, 15번은 다시 조정할 필요성이 있다. 또한 K-척도는 4개의 요인으로 분석했는데 설문지를 분석한 결과 전체 4개 요인을 사용하는 것보다는 영향력 있는 2개의 요인을 이용하여 집단들을 분류하는 것이 더욱 효과적이었다. 자가진단 프로그램이므로 서로 관련 있는 많은 문항보다는 대표성 있는 적은 문항으로 빠른 시간 내에 진단할 수 있다면 많은 사람들이 자가진단을 통해 자신의 상태를 파악할 수 있을 것이다.

참고문헌

- Choi, H. S. and Ha, J. (2011). A study on the factors inducing internet addiction of college students. *Journal of the Korean Data & Information Science Society*, **22**, 437-448.
- Gjorgji, M., Dejan, G. and Ivan, C. (2009). A multi-class svm classifier utilizing binary detection tree. *Informatica*, **33**, 233-241.

- Hwang, C. (2011). Asymmetric least squares regression estimation using weighted least squares support vector machine. *Journal of the Korean Data & Information Science Society*, **22**, 999-1005.
- Hwang, C. (2010). Variable selection for multiclassification by LS-SVM. *Journal of the Korean Data & Information Science Society*, **21**, 959-965.
- Jung, Y. T., Oh, J. A. and Lee, Y. J. (2012). The study on factors affecting committing behavior school violence according to the degree of internet addiction. *Journal of Digital Policy*, **10**, 77-86.
- Kim, J. W. and Kim, H. J. (2011). Factor analysis for improving adults internet addiction diagnosis. *International Journal of Fuzzy Logic and Intelligent systems*, **15**, 317-322.
- Kim, H. J. and Kim, J. W. (2012). Comparative study of K-scale and the internet addiction diagnosis method using tolerance degree for internet use. *Korea Association of Computer Education*, **15**, 47-55.
- Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium*, Berkeley, University of California Press, 481-492.
- Lee, J. and Lee, D. (2005). An improved cluster labeling method for support vector clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 461-464.
- Mercer, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society A*, 415-446.
- Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A. J. and Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, **13**, 1443-1472.
- Shim, J. and Hwang, C. (2011). Forecasting volatility via conditional autoregressive value at risk model based on support vector quantile regression. *Journal of the Korean Data & Information Science Society*, **22**, 589-596.
- Suykens, J. A. K. (2000). Least squares support vector machine for classification and nonlinear modeling. *Neural Network World, Special Issue on PASE 2000*, **10**, 29-48.
- Suykens, J. A. K. and Vanderwalle, J. (1999). Least square support vector machine classifier. *Neural Processing Letters*, **9**, 293-300.
- Suykens, J. A. K. and Vanderwalle, J. and De Moor, B. (2001). Optimal control by least squares support vector machines. *Neural Networks*, **14**, 23-35.
- Tax, D. M. J. and Duin, R. P. W. (1999). Support vector domain description. *Pattern recognition Letters*, **20**, 1191-1199.
- Vapnik, V. (1995). *The nature of statistical learning theory*, Springer-Verlag, New York.
- Vapnik, V. (1998). *Statistical learning theory*, Springer, New York.
- Yang, J., Estivill-Castro, V. and Chalup, S. K. (2002). Support vector clustering through proximity graph modelling. *Proceedings of the 9th International Conference. Neural Information Processing*, **2**, 898-903.

Study on the K-scale reflecting the confidence of survey responses

Hye Jung Park¹ · Su Young Pi²

¹Daegu University

²Faculty of Liberal Education, Catholic University of Daegu

Received 12 November 2012, revised 17 December 2012, accepted 24 December 2012

Abstract

In the Information age, internet addiction has been a big issue in a modern society. The adverse effects of the internet addiction have been increasing at an exponential speed. Along with a great variety of internet-connected device supplies, K-scale diagnostic criteria have been used for the internet addiction self-diagnose tests in the high-speed wireless Internet service, netbooks, and smart phones, etc. The K-scale diagnostic criteria needed to be changed to meet the changing times, and the diagnostic criteria of K-scale was changed in March, 2012. In this paper, we analyze the internet addiction and K-scale features on the actual condition of Gyeongbuk collegiate areas using the revised K-scale diagnostic criteria in 2012. The diagnostic method on internet addiction is measured by the respondents' subjective estimation. Willful error of the respondents can be occurred to hide their truth. In this paper, we add the survey response to the trusted reliability values to reduce response errors on the K-scale on the K-scale, and enhance the reliability of the analysis.

Keywords: Cluster analysis, factor analysis, Gaussian kernel, internet addiction, K-scale, support vector cluster.

¹ Corresponding author: Invitation professor, Daegu University, Gyeongbuk 712-714, Korea.
E-mail: hyjpark@daegu.ac.kr

² Assistant professor, Faculty of Liberal Education, Catholic University of Daegu, Gyeongbuk 712-702, Korea.