

## 성인 인터넷 중독진단 개선을 위한 요인분석

# Factor Analysis for Improving Adults' Internet Addiction Diagnosis

김종완 · 김희재

Jongwan Kim and Hee-Jae Kim

대구대학교 컴퓨터 · IT공학부

Division of Computer and Information Technology, Daegu University

### 요 약

한국정보화진흥원에서 개발한 한국형 성인 인터넷 중독 자가진단 척도인 K-척도는 4가지 요인의 20 문항으로 구성되어 있으며, 사용자의 설문응답값으로 인터넷 중독을 진단한다. 기존의 연구는 대부분 인터넷 중독의 원인을 찾으려는 시도였으며, 청소년 대상으로 수집된 표본을 가지고 그들의 인터넷 중독진단이 수행되었다. 본 연구의 목적은 통계 기법의 주성분분석과 데이터마이닝 기법인 의사결정트리를 이용하여 K-척도의 사용자군 분류를 판정하는 주요인을 발견하는 것이다. 실험 결과로부터 K-척도를 구성하는 4가지 요인 중 내성 및 몰입 요인이 성인 인터넷 중독진단에 가장 큰 영향을 주는 요인임을 알 수 있었다.

**키워드** : 인터넷 중독진단, K-척도, 요인분석, 의사결정트리, 주성분분석

### Abstract

Korean adults' internet addiction diagnosis measure, K-scale developed by Korea National Information Society Agency (NIA), has composed of 4 categories including 20 items. This scale can diagnose user's internet addiction with individual's questionnaire items. Most of previous research works were tried to know reasons of internet addiction and to judge whether adolescents are addicted or not with their samples. In this research, it is the goal to find the key component to judge individual's internet addiction by using a decision tree in the data mining field and a principal component analysis in statistics. From the experimental results, we would discover that tolerance and preoccupation factor is the most important one to affect adult's internet addiction.

**Key Words** : Internet Addiction Diagnosis, K-scale, Factor Analysis, Decision Tree, Principal Component Analysis.

## 1. 서 론

인터넷 중독(internet addiction)은 '정보 이용자가 지나치게 컴퓨터에 접속하여 일상생활에 심각한 육체적, 정신적, 사회적 그리고 금전적인 지장을 받고 있는 상태'라고 정의하기도 하고, 도박중독, 알코올중독, 마약중독 등과 비슷한 현상으로 지나친 인터넷 접속으로 인하여 정보 이용자가 정서 불안정이나 강박적 집착으로 스스로 통제가 불가능하며, 내성, 의존성, 금단증상 등이 발생하여 일상생활이 힘들어진 상태를 말하기도 한다[1]. 해마다 전 국민의 인터넷 의존도와 인터넷 이용률이 향상되어, 스스로의 통제가 부족한 유아, 청소년뿐만 아니라 스스로의 통제가 가능한 성인임에도 불구하고 실업과 인터넷 역기능의 증가에 맞물려 매년 증가하는 인터넷 중독율은 큰 사회문제가 되고 있다.

Kimberly Young에 의해 최초로 개발된 인터넷 중독진단 도구인 Young척도는 인터넷 중독을 알콜중독처럼 하나의 병리학적 중독과 같은 특성으로 보고, 인터넷 중독을 5가지 유형 즉, 사이버 성적 중독, 사이버 관계 중독, 넷 강박충동, 정보 과부하, 컴퓨터 게임 중독으로 나누어, 5점 척도 20개의 설문문항으로 인터넷 중독진단을 시행하고 있다[2]. 이를 바탕으로 한국정보화진흥원에서 개발한 K-척도는 인터넷을 통한 가상세계지향(6문항), 긍정적 기대(4문항), 내성 및 몰입(6문항), 인터넷에 대한 자기인식(4문항)에 대한 4요인, 총 20개의 설문문항으로 사용자의 인터넷 중독 자가진단을 하고 있다[3]. 국내외에서 인터넷 중독의 원인 및 진단과 치료에 대한 연구가 활발하게 진행되고 있다[4-7].

2005년에 Chou 등은 인터넷 중독의 사회적 효과에 대하여 기존의 다양한 연구를 주제별로 정리하고, 현재 및 미래의 인터넷 중독 치료방안과 향후 연구주제들을 제시하였다[4]. 국내에서는 2008년 기준으로 국내 학술지에 게재된 관련 연구 논문 173편을 중심으로 수행한 인터넷 중독 연구동향 분석 연구가 있다[5]. 권재환씨가 수행한 연구내용과 연구대상 등의 평가기준으로 분석한 결과에 따르면, 대부분의 연구가 중고등학생 대상으로 이루어졌으며, 연구내용은 주로 인터넷 중독의 심리, 사회, 환경적 원인과 결과를 찾는

접수일자 : 2011년 3월 19일

완료일자 : 2011년 5월 26일

본 논문은 본 학회 2011년도 춘계 학술대회에서 선정된 우수논문입니다.

이 논문은 2010학년도 대구대학교 학술연구비 지원에 의하여 연구되었음.

내용이나 인터넷 중독과 관련된 변인들 사이의 관계를 다루는 유형이 전체의 80% 수준을 차지하고 있음을 보여준다. 기술적으로 인터넷 중독의 진단과 처방을 위한 전문가 시스템 설계 및 구현에 관한 연구[6]에서는 사용자의 인터넷 사용 데이터를 자동으로 수집하고, 인간 전문가의 치료 지식을 활용하여 인터넷 중독에 대한 자동진단 및 처방이 가능한 전문가 시스템을 Java 언어 기반의 JESS 툴을 사용하여 구현하였다. 인간 전문가의 처방에 약 90% 정도 일치하는 성능을 실험결과에서 보여주고 있으며, 자동으로 인터넷 중독진단과 처방 방법을 찾으려는 첫 번째 시도로 보인다. 최근 중국의 청소년 대상으로 뇌파 측정 장치를 이용한 연구 [7]에서는 심리 발달 측면에서 중국 청소년들의 인터넷 중독을 분류하여 진단방법을 제시하고, 인과관계 분석 등을 수행한 바 있다.

본 연구에서는 4점 척도, 20개의 설문문항으로 모든 사용자에게 동일하게 측정되는 K-척도를 대학생과 일반인들로부터 구성된 669명에게 실시하였다. 그 중 불성실하게 응답한 설문지 43부를 제외하고 성실하게 기재된 626명의 K-척도 설문응답값을 보편적인 통계적 기법에 사용되는 SPSS와 데이터마이닝 도구로 많이 사용되는 WEKA에 구동하여 요인분석을 수행하고 그 결과를 제시한다.

이 논문은 다음과 같이 구성된다. 2장에서는 성인 인터넷 중독의 자가진단 척도인 K-척도를 소개하고, 3장에서는 SPSS를 활용한 요인분석 결과와 WEKA를 활용한 결과를 각각 제시한다. 결론은 4장에 기술된다.

## 2. 성인 인터넷 중독 자가진단 척도

한국정보화진흥원에서는 2006년 20문항으로 이루어진 한국형 성인 인터넷 중독 자가진단 척도(K-척도)를 개발하여 일상생활 장애, 급단, 내성 및 몰입, 일탈행동 등 증상에 따라 집중 치료를 요망하는 인터넷 중독 고위험사용자군, 상담을 요망하는 잠재적위험사용자A군, 자기 관리를 요망하는 잠재적위험사용자B군, 건전한 사용자인 일반사용자군으로 분류하여 한국인에 맞게 스스로 인터넷 중독진단 및 상담 치료를 할 수 있도록 웹서비스하고 있다[3]. K-척도의 진단결과는 아래 그림 1과 같이 사용자군의 판별점수는 설문응답자의 응답값의 총합에 따라 4개의 사용자군으로 분류된다[3].

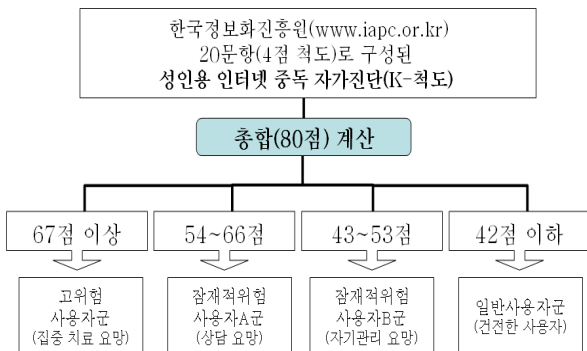


그림 1. K-척도 진단방법

Fig. 1. K-scale diagnostic method

그림 1의 K-척도 진단방법을 이용하여 설문응답자 626명을 대상으로 인터넷 중독을 진단한 결과는 아래 표 1과

같다. 설문응답자들 가운데 총합 80점에 대하여 인터넷 중독 자가진단 응답값이 67점 이상 사용자인 고위험사용자군은 IT전공학생 2명밖에 없는 것으로 진단되었다. 이것은 실제보다 자신의 중독정도를 낮추려는 심리적 요인으로 보이며, 차후 연구에서 더 밝혀야 할 부분이다.

표 1. 사용자별 K-척도 진단결과

Table 1. The diagnostic result of K-scale for each user

구분	일반 사용자군	잠재적위험 사용자B군	잠재적위험 사용자A군	고위험 사용자군
IT전공학생 (170명)	150 88.24%	13 7.65%	5 2.94%	2 1.18%
IT비전공학생 (178명)	161 90.45%	15 8.43%	2 1.12%	0 0.00%
IT전공일반 (104명)	97 93.27%	5 4.81%	2 1.92%	0 0.00%
IT비전공일반 (174명)	161 92.53%	10 5.75%	3 1.72%	0 0.00%
전체 (626명)	569 90.89%	43 6.87%	12 1.92%	2 0.32%

한편, K-척도는 앞서도 설명했듯이, 4가지 요인 즉, 6문항의 가상세계지향요인과 4문항의 긍정적 기대요인, 6문항의 내성 및 몰입요인과 4문항의 인터넷에 대한 자기인식요인으로 구성된다. 본 연구에서는 설문참여자 626명을 대상으로 K-척도 4가지 요인 가운데 어느 요인이 사용자군 판별에 가장 큰 영향을 주는지 알아보기 위해서, K-척도의 각 요인별 총합을 구하여 통계 기법의 요인분석과 데이터마이닝 기법의 의사결정트리를 적용하려고 한다.

## 3. 인터넷 중독진단을 결정하는 요인분석

본 연구에서는 설문응답자 626명을 대상으로 20문항의 K-척도 설문조사를 실시하여 구한 설문응답값을 가지고, 인터넷 중독진단을 결정하는 요인분석을 하고자 통계적 기법과 데이터마이닝 기법을 사용하였다.

### 3.1 통계적 방법을 활용한 요인분석

통계적 기법에서는 다수 변수들 사이의 관계를 분석하여 변수들의 공통 속성을 통해 변수들을 설명하기 위한 방법으로 요인분석(factor analysis)을 사용한다. 요인분석 기법에는 주성분분석(PCA: Principal Component Analysis)과 공통요인분석(CFA: Common Factor Analysis)이 있다. 주성분분석은 데이터 집합을 분석하는 방법 가운데 하나로서, 데이터를 한 개의 축으로 사상시켰을 때 그 분산이 가장 커지는 축이 첫 번째 좌표축으로 오고, 두 번째로 커지는 축이 두 번째로 위치시키는 방식으로 차례로 놓이도록 새로운 좌표계로 데이터를 선형 변환한다. 이와 같이 각각의 축에 데이터의 가장 중요한 성분을 위치시킴으로써 여러 가지 응용이 가능하다[8].

본 연구에서는 대표적인 통계적 기법인 SPSS의 요인분석 메뉴에 있는 주성분분석을 이용하여 6문항 또는 4문항으로

구성된 4가지 요인 중 어느 요인이 K-척도 진단결과에 큰 영향을 미치는지에 대해 알아본다. 주성분분석 중 적절한 요인 변수의 개수를 찾는 성분행렬과 스크리 도표를 이용하여, K-척도는 몇 개의 주된 성분으로 구성되었는지 밝히고자 한다.

먼저, 실험 참여자 626명에 대한 K-척도 20문항의 설문 응답값을 이용하여 4가지 요인에 대한 기술통계량을 구한 결과는 아래 표 2와 같다.

표 2. K-척도의 요인별 기술통계량  
Table 2. Technical statistics for each factor of K-scale

요인	평균	분산	범위	중위수	최빈값	왜도
1.가상세계지향(6문항)	1.41	0.15	6~21	8	8	1.76
2.긍정적기대(4문항)	1.48	0.26	4~16	5	4	1.36
3.내성및몰입(6문항)	1.90	0.34	6~24	11	10	0.45
4.인터넷에대한자기인식(4문항)	1.42	0.28	4~16	5	4	1.52

각 사용자의 요인별 합계를 구한 후 실험 참여자 전체에 대한 각 요인별 평균과 분산 및 기술통계량을 구한 표 2를 통하여, 요인 3인 '내성 및 몰입'요인의 기술통계량은 다른 요인들에 비해 절대적 차이는 크지 않지만 상대적으로 값이 크다는 것을 알 수 있다. 즉 다른 세 가지 요인 1, 2, 4의 평균값은 약 1.4로 비슷한데 비하여, 요인 3의 평균값은 1.9이므로 평균적으로 약 32% 정도의 차이로 확연히 구분된다. 또한 평균을 중심으로 데이터들의 흩어진 정도를 나타내는 분산을 보면, 요인 3은 요인 1, 2, 4의 분산에 비해 평균적으로 약 60% 정도의 차이로 다른 요인들보다 자료들이 넓게 분포되어 있음을 알 수 있다. 뿐만 아니라 4점 척도인 K-척도에서 중위수와 최빈값이 동일한 요인 2 및 요인 4와는 달리, 6문항으로 구성된 요인 1은 설문응답자 626명으로부터 6문항 중 4의 척도값을 모두 선택한 사용자가 없지만 요인 3은 626명 중 4의 척도값을 모두 선택한 사용자가 있어서 범위와 중위수 및 최빈값이 요인 1에 비해 크다는 것을 알 수 있다.

왜도(skewness)는 분포의 모양이 어느 쪽으로 치우쳐 있는가를 나타내는 척도로 완전히 대칭분포를 이루는 경우 왜도 값이 0이 된다[9]. 반면 왼쪽 비대칭성 분포일 경우에는 음의 왜도 값으로 중위수를 포함한 데이터가 오른쪽에 더 많이 분포되어 분포의 모양이 왼쪽 부분에 긴 꼬리를 가지게 되며, 양의 왜도 값이면 데이터가 왼쪽에 더 많이 분포되어 오른쪽 부분에 긴 꼬리를 가지는 오른쪽 비대칭성분포가 된다[9]. 표 2의 요인 3은 왜도가 0.45로 평균과 중위수의 주위에 대부분의 데이터가 분포되어 대칭모양을 이루고 있는 그림 2의 (c)이며, 다른 요인들은 왜도가 1.3이상의 양의 값으로 데이터들이 왼쪽으로 치우쳐 있으며 오른쪽 부분에 긴 꼬리를 가지는 비대칭성 분포인 그림 2의 (a), (b), (d)임을 알 수 있다.

비록 K-척도는 설문응답자의 주관적인 판단의 편중이 있을 수 있지만, 요인 3에 대한 설문응답값의 분포가 고르게 분포되어 있다는 것은 '가상세계지향', '긍정적 기대', '인터넷에 대한 자기인식'의 요인과는 달리 '내성 및 몰입'요인이 모든 인터넷 사용자들이 인터넷을 이용하는데 가지는 공통 성향이라는 것을 확인할 수 있다.

한편 K-척도 4가지 요인에 대하여 추출될 주성분의 수를 구하는 방법에는 고유값(eigenvalue) 이용과 전체 성분들의 설명력 기준과 스크리 도표(scree plot) 등의 방법이 사용된다[10]. 고유값은 해당 성분이 설명하는 분산의 양을 나타내는데, 값이 클수록 중요한 성분이라 판단할 수 있으며, 1이상의 고유값을 갖는 성분의 수만큼 주성분의 수를 추출한다. 그림 3의 K-척도 4가지 요인에 대한 스크리 도

표는 공통된 4가지 성분의 고유값을 보여주는데, 첫 번째 성분에서 마지막 네 번째 성분으로 갈수록 고유값은 줄어 들고 있으며 1이상의 고유값을 가지는 성분은 1개임을 알 수 있다. 이것은 4가지 요인들이 공통되는 하나의 주된 성분을 가지고 있음을 의미한다.

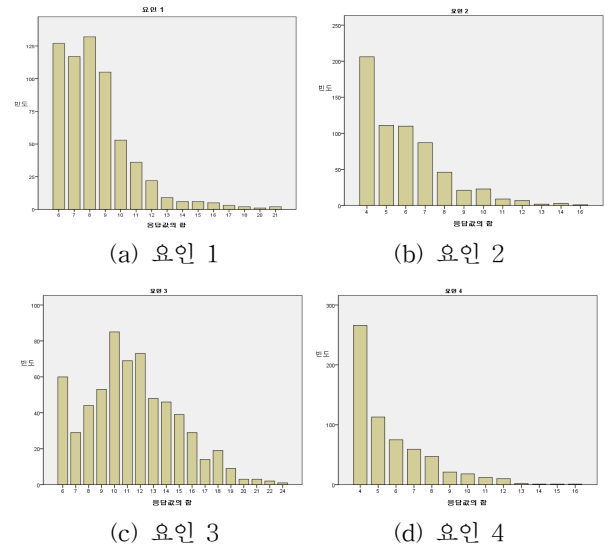


그림 2. 각 요인들의 왜도  
Fig. 2. Skewness of each factor

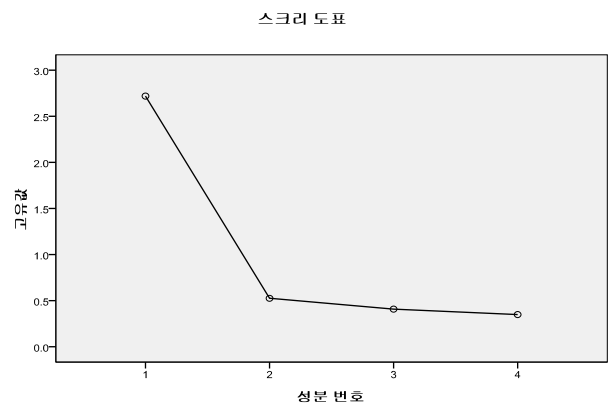


그림 3. K-척도 4요인에 대한 스크리 도표  
Fig. 3. Scree plot for 4 factors of K-scale

다음으로 추출할 성분 개수를 나타내는 표 3의 성분행렬(component matrix)을 통하여 4가지 요인이 모두 하나의 성분으로 성분 1에 나타나는 것을 알 수 있다. 이것은 1이상의 고유값이 1개인 그림 3의 스크리 도표의 결과와도 일치한다.

표 3. K-척도의 성분행렬  
Table 3. Component matrix of K-scale

요인	성분 1
요인 3	0.852
요인 1	0.839
요인 4	0.813
요인 2	0.793

또한 0과 1사이의 값을 갖는 공통성(communality)을 이용하여 추출된 성분에 의해 각 요인이 얼마나 설명되는지 알아본다. 공통성이란 요인의 분산이 추출된 성분들에 의해 설명되는 정도를 나타내며, 0.5를 넘어야 의미가 있다[10]. 표 4의 결과로부터 모든 요인의 공통성이 0.5이상으로 공통성에 대한 분석이 가능하며, 추출된 성분에 의하여 모든 요인들은 약 63%~73%가 설명되어짐을 알 수 있다. 특히 70%이하를 나타내는 다른 요인들보다 요인 3은 추출된 성분에 의해 72.5%가 설명됨을 확인할 수 있다.

표 4. K-척도의 공통성  
Table 4. Communality of K-scale

요인	초기값	추출값
요인 1	1.000	0.704
요인 2	1.000	0.629
요인 3	1.000	0.725
요인 4	1.000	0.661

마지막으로 4가지 요인들로부터 추출된 4개의 성분에 대하여 각 성분의 고유값과 각 성분의 설명력을 보여주는 ‘% 분산’이 계산된 표 5의 내용은 다음과 같다. 각 성분의 고유값을 나타내는 ‘합계’를 통하여, 1이상의 고유값을 가지는 주성분이 1개임을 알 수 있으며, 각 성분이 전체 분산의 몇 %를 설명하는 가를 나타내는 ‘%분산’을 통하여, 추출된 하나의 주성분이 전체 분산의 약 68%를 설명한다는 것을 알 수 있다.

표 5. 성분별 고유값과 설명력  
Table 5. Eigenvalues and their explainabilities for each component

성분	초기 고유값		
	합계	% 분산	% 누적
1	2.719	67.969	67.969
2	0.525	13.124	81.093
3	0.408	10.193	91.286
4	0.349	8.714	100.000

결론적으로 K-척도의 4가지 요인에 대한 통계적 기법의 주성분분석을 사용하여, 추출된 하나의 성분은 K-척도 전체 분산의 약 68%를 설명하며, 요인 3은 추출된 주성분에 의해 72.5%가 설명되는 가장 중요한 요인임을 알 수 있다.

3.2 WEKA를 활용한 요인분석

데이터마이닝은 대용량의 데이터로부터 데이터 내에 존재하는 관계, 패턴, 규칙 등을 탐색하고 찾아내어 모형화하므로써 유용한 지식을 추출하는 일련의 과정을 의미한다[11]. 의사결정트리(decision tree)는 데이터마이닝의 의사결정규칙을 트리구조로 도표화하여 분류와 예측을 수행하는 분석방법으로, 분류 또는 예측 과정이 트리구조에 의한 추론규칙(induction rule)으로 표현되기 때문에 과정이 쉽게 이해되고 설명할 수 있는 장점이 있다.

본 연구에서는 설문 참여자 626명의 K-척도 각 요인별 설문응답값을 독립변수로, K-척도 진단방법에 따라 분류된 각 사용자군 값을 종속변수로 설정하여, 의사결정트리 분석을 위해 C4.5알고리즘을 WEKA[11]에서 구현한 J48을 이

용하였다. 또한, WEKA의 Select attributes 메뉴의 Information Gain 기능을 이용하여 K-척도 4가지 요인 중 어느 요인이 사용자군 분류에 가장 큰 영향이 주는지 알아 보았다.

먼저, 실험 참여자의 K-척도에 따른 인터넷 중독 사용자군 분류에 대한 의사결정트리를 발견하기 위하여 WEKA의 J48을 사용하여 K-척도 4가지 요인에 대한 요인분석을 수행한 결과는 그림 4와 같다[12]. 그림 4에 주어진 의사결정트리의 루트노드에 요인 3이 위치하므로, 요인 3이 인터넷 중독진단의 각 사용자군을 분류하는 핵심 요인임을 알 수 있다. 그림 4의 트리구조에 대한 이해를 돕기 위하여 분류 규칙 형태로 바꾸어서 표현한 그림 5를 해석하면, 요인 3(factor 3)이 15이하이고 요인 1이 11이하인 설문응답자는 일반사용자군(GU)으로 분류된다. 또한 요인 3이 15이하이고 요인 1이 11을 초과하고 요인 4가 8을 초과하면 잠재적 위험사용자B군(PAB)이 된다. 나머지 규칙들도 이와 같이 정리하면, 일반사용자군은 규칙 7개, 잠재적위험사용자A군은 규칙 2개, 잠재적위험사용자B군은 규칙 8개, 고위험사용자군은 규칙이 1개로, 모두 18개의 분류규칙이 구해진다.



그림 4. 실험 참여자의 J48에 의한 의사결정트리  
Fig. 4. Decision tree generated by J48 for overall participants

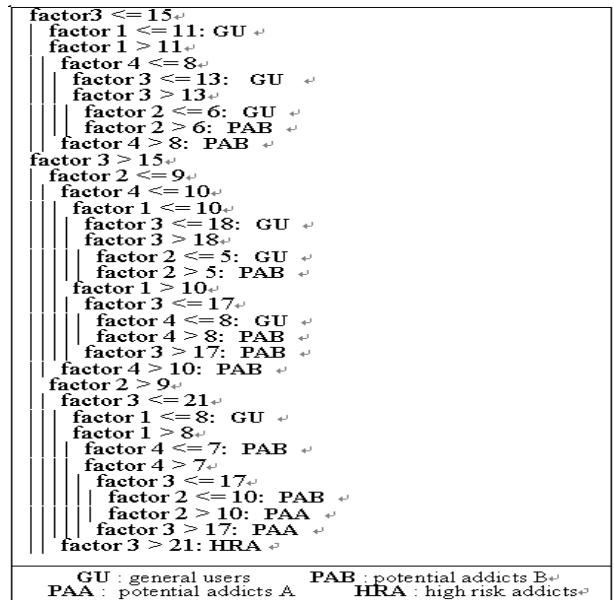


그림 5. 실험 참여자의 의사결정 분류규칙  
Fig. 5. Decision classification rules generated by J48 for overall participants

다음 표 6은 실험 참여자의 의사결정트리 분류규칙 18개를 정리한 것으로 각 규칙들의 논리곱인 ‘and’를 기호 ‘&’

로 표현하였다. 설문응답자의 K-척도 응답값 총합에 따라 분류되는 규칙을 통해 각 요인의 값에 따라 사용자군을 쉽게 분류할 수 있다. 18개의 모든 규칙에 요인 1은 2번, 요인 2와 요인 4는 3번 생략되어 있지만, 요인 3은 모두 포함되어 있어, 분류규칙면에서도 요인 3이 중요한 요인이라는 것을 확인할 수 있다.

표 6. 실험 참여자의 의사결정트리 분류규칙  
Table 6. Classification rules of a decision tree for overall participants

번호	규칙
1	if 요인1≤11 & 요인3≤15 then 일반사용자군
2	if 요인1>11 & 요인3≤13 & 요인4≤8 then 일반사용자군
3	if 요인1>11 & 요인2≤6 & 13<요인3≤15 & 요인4≤8 then 일반사용자군
4	if 요인1>11 & 요인2>6 & 13<요인3≤15 & 요인4≤8 then 잠재적위험사용자B군
5	if 요인1>11 & 요인3≤15 & 요인4>8 then 잠재적위험사용자B군
6	if 요인1≤10 & 요인2≤9 & 15<요인3≤18 & 요인4≤10 then 일반사용자군
7	if 요인1≤10 & 요인2≤5 & 요인3>18 & 요인4≤10 then 일반사용자군
8	if 요인1≤10 & 5<요인2≤9 & 요인3>18 & 요인4≤10 then 잠재적위험사용자B군
9	if 요인1>10 & 요인2≤9 & 15<요인3≤17 & 요인4≤8 then 일반사용자군
10	if 요인1>10 & 요인2≤9 & 15<요인3≤17 & 8<요인4≤10 then 잠재적위험사용자B군
11	if 요인1>10 & 요인2≤9 & 요인3>17 & 요인4≤10 then 잠재적위험사용자B군
12	if 요인2≤9 & 요인3>15 & 요인4>10 then 잠재적위험사용자B군
13	if 요인1≤8 & 요인2>9 & 15<요인3≤21 then 일반사용자군
14	if 요인1>8 & 요인2>9 & 15<요인3≤21 & 요인4≤7 then 잠재적위험사용자B군
15	if 요인1>8 & 요인2=10 & 15<요인3≤17 & 요인4>7 then 잠재적위험사용자B군
16	if 요인1>8 & 요인2>10 & 15<요인3≤17 & 요인4>7 then 잠재적위험사용자A군
17	if 요인1>8 & 요인2>9 & 17<요인3≤21 & 요인4>7 then 잠재적위험사용자A군
18	if 요인2>9 & 요인3>21 then 고위험사용자군

그 결과 표 7과 같이 설문 참여자의 의사결정트리 분류 규칙에 따른 각 사용자군 분류와 K-척도 사용자군 분류와의 교차분석을 통한 일치율은 99.2%(= {569+38+12+2}/626=0.992)이며, K-척도시 잠재적위험사용자B군으로 분류되었지만 전체 사용자의 의사결정트리 분류규칙에서는 4명의 일반사용자군과 1명의 잠재적위험사용자A군으로 분류된 5명의 불일치율은 0.8%(= {4+1}/626=0.008)로 나타났다. 이 결과로부터 K-척도와 의사결정트리 분류규칙간에는 99%

수준의 일치도를 보이므로 신뢰도가 높다고 판단된다.

따라서 설문 참여자의 K-척도 의사결정트리를 나타내는 그림 4와 그림 5로부터 루트노드가 요인 3이므로, 요인 3이 다른 요인보다 K-척도의 각 사용자군 분류에 큰 영향을 끼친다는 사실이 확인된다. 또한 의사결정트리 분류규칙을 통하여 K-척도에서 잠재적위험사용자B군에 속하지만, K-척도 진단보다 위험성이 있는 표출되지 않은 1명의 잠재적위험사용자A군의 존재를 알려준다. 특히 K-척도의 의사결정트리는 트리구조의 추론규칙을 가지고 논리적으로 각 사용자군 분류를 설명할 수 있으므로 사용자를 이해시키는데 도움이 된다는 장점이 있다.

표 7. 실험 참여자의 의사결정트리 분류규칙과 K-척도의 진단결과 비교

Table 7. Comparison of diagnostic results of classification rules of decision tree and them of K-scale for overall participants

전체 사용자 구분	전체 사용자의 의사결정트리 분류규칙 적용				합계	
	일반 사용자군	잠재적위험 사용자B군	잠재적위험 사용자A군	고위험 사용자군		
K-척도	일반 사용자군	569	0	0	0	569
	잠재적위험 사용자B군	4	38	1	0	43
	잠재적위험 사용자A군	0	0	12	0	12
	고위험 사용자군	0	0	0	2	2
합계	573	38	13	2	626	

부가적으로 WEKA의 Select attributes메뉴에 있는 Information Gain(IG) 기능을 이용하여 K-척도 4가지 요인 중 어떤 요인이 각 사용자군 분류에 가장 큰 영향을 주는지 확인한다. IG는 기계학습 분야에서 자주 사용되는 속성 선택(feature selection) 기법으로 데이터에서 속성의 출현 빈도뿐만 아니라 출현하지 않은 빈도까지 고려해서 각 범주에서의 속성 정보량을 계산하는 방법이다[11].

표 8. 실험 참여자의 K-척도 요인별 Information Gain  
Table 8. Information Gain of each factor in K-scale for overall participants

요인	요인 1	요인 2	요인 3	요인 4
IG	0.235	0.209	0.249	0.241

IG는 데이터를 잘 구분하기 위하여 어떤 속성을 선택해야 되는지를 나타내는 값으로, 값이 클수록 데이터를 잘 구분한다고 할 수 있다. 표 8은 K-척도의 IG를 구한 결과로 3.2%~16.1%의 차이가 있으며, 요인 3이 다른 요인들에 비해 인터넷 중독에 대한 각 사용자군 분류를 잘하는 것을 보여준다. 이것은 통계 기법인 SPSS의 요인분석 결과와 데이터마이닝 기법이 일치함을 나타낸다.

#### 4. 결 론

본 연구에서는 한국형 성인 인터넷 중독 자가진단 척도인 K-척도를 구성하는 4가지 요인 중 어느 요인이 인터넷 중독진단에 대한 각 사용자군 분류의 핵심 요인인지를 밝히



고자 하였다.

이를 위해서 통계적 분석방법인 SPSS를 이용하여 K-척도를 구성하는 각 요인들의 분산과 범위 등과 같은 기술통계량을 살펴본 결과, 다른 요인들과는 달리 대부분의 사용자가 요인 3인 내성 및 몰입 성향을 갖고 있음을 알 수 있었다. 또한 주성분분석(PCA) 수행으로, 내성 및 몰입 요인이 인터넷 중독진단을 결정하는 주성분임을 발견할 수 있었다.

또한 대표적인 데이터마이닝 도구인 WEKA를 이용하여 인터넷 중독에 대한 4가지 사용자군을 분류함으로써 모두 18개의 의사결정트리 분류규칙을 발견했으며, 이로 인해 타사용자군에 잠재된 사용자군들을 찾을 수 있었다. 또한 K-척도 4가지 요인의 중요도를 발견하기 위한 IG 계산을 통하여, 요인 3이 다른 요인에 비해 중요한 요인임을 알 수 있었다.

결론적으로 본 연구를 수행하여 발견한 두 가지 중요한 사실은 다음과 같다. 첫째, 요인 3인 내성 및 몰입 요인이 인터넷 중독진단을 결정하는 핵심 요인이라는 사실이다. 둘째, 두 가지 방법의 분석결과로부터 인터넷 사용자들은 자신의 중독성을 간과하는 경향이 있으며, K-척도와 지능적 분석방법 간에는 일부 차이가 있음도 알 수 있었다. 이에 본 연구진은 다른 3가지 요인보다 요인 3에 가중치를 부여하는 새로운 인터넷 중독 설문문항을 개발하고 이를 이용한 인터넷 중독진단 방법을 연구하고 있다[13].

향후에는 전체 설문응답자 대상이 아닌 IT전공학생, IT비전공학생, IT전공일반인, IT비전공일반인의 각 사용자 직무에 따라 의사결정트리를 구한 후, K-척도와외의 불일치도를 비교하여 보다 정확한 인터넷 중독진단을 수행하고자 하며, 현재의 연구결과를 검증하기 위해서 보다 구체적인 실험을 통하여 성능도 평가하고자 한다.

### 참 고 문 헌

[1] 인터넷윤리실천협의회, 한국정보처리학회, *U시대의 인터넷 윤리*, 2판, 이한출판사, 2009.

[2] K. Young, "Internet addiction: The emergence of a new clinical disorder", *Cyber-Psychology and Behavior*, vol. 1, no. 3, pp. 237-249, 1996.

[3] 한국정보화진흥원, 인터넷 중독 예방상담센터, <http://iipc.or.kr/>.

[4] C. Chien, C. Linda, and B. John, "A Review of the Research on Internet addiction", *Educational Psychology Review*, vol.17, no.4, pp.363-388, 2005.

[5] 권재환, "국내 인터넷 중독 연구동향 - 학술지 게재 논문 분석", *청소년학연구*, 제15권, 3호, pp. 137-157, 2008.

[6] 송원문, 이상화, 김은주, 송성렬, 송수민, 김명원, "인터넷 중독의 진단과 처방을 위한 전문가 시스템", *한국지능시스템학회 춘계학술대회논문집*, 제20권, 1호, pp. 289-292, 2010.

[7] W. Gao, F. Zhang, and Z. Chen, "Psychopathological Mechanism and Compensation Psychotherapy of Internet addiction", *In Proc. of the first IEEE Symposium on Web Society*, pp. 172-176, 2009.

[8] R. Gorsuch, *Factor Analysis*, Lawrence Erlbaum, 1983.

[9] 김희탁, 유한주, 황복주, *EXCEL을 활용한 통계학*, 제2판, 법문사, 2008.

[10] 이학식, 임지훈, *SPSS 14.0 매뉴얼*, 법문사, 2008.

[11] I. Witten and E. Frank, *Data Mining: practical machine learning tools and techniques*, 2nd, Morgan Kaufmann, 2005.

[12] 김종완, 김희재, "성인 인터넷 중독진단을 위한 요인분석", *한국지능시스템학회 춘계학술대회논문집*, vol. 21, no. 1, pp. 212-213, 2011.

[13] 김희재, 김종완, "인터넷 주활동 분석을 통한 사용자의 인터넷 중독진단에 관한 연구," *한국산업정보학회논문지*, 계재승인, 2011.

### 저 자 소 개



#### 김종완(Jongwan Kim)

1987년 : 서울대학교 컴퓨터공학과(공학사)  
 1989년 : 서울대학교 컴퓨터공학과(공학석사)  
 1994년 : 서울대학교 컴퓨터공학과 졸업  
 (공학박사)  
 1995년~현재 : 대구대학교 컴퓨터·IT공학부 교수  
 1999년~2000년 : 미국 U. of Massachusetts 방문교수  
 2006년-2007년 : 미국 U. of Oregon 객원교수

관심분야 : 인공지능, 데이터마이닝, 개인화, IT융합서비스, 인터넷 역기능  
 Phone : 053-850-6575  
 Fax : 053-850-6589  
 E-mail : jwkim@daegu.ac.kr



#### 김희재(Hee Jae Kim)

1992년 : 대구가톨릭대학교 통계학과(학사)  
 1994년 : 대구가톨릭대학교 전산통계학과  
 (이학석사)  
 2005년 : 대구대학교 컴퓨터정보공학과 박사 수료  
 2005년~현재 : 대구대학교 컴퓨터·IT공학부 겸임교수

관심분야 : 인터넷 중독, 퍼지시스템, 인공지능, 데이터마이닝  
 E-mail : heejae0305@daegu.ac.kr