

청소년의 컴퓨터 오락추구 행동을 예측하기 위한 신경망 활용

이혜주[†] · 정의현^{††}

요 약

본 연구에서는 신경망을 활용하여 청소년의 컴퓨터 오락추구 행동을 설명하는 예측모형을 조사하고자 하였다. 이를 위해 한국청소년 패널 조사(KYPS)의 중 2패널의 1차년도 데이터(총 3449명, 남: 1725명, 여: 1724명)를 대상으로 하여 신경망 모형(모형 1)을 구축하였다. 또한 신경망 모형의 성능을 분석하고자 로지스틱 회귀 분석을 실시하고 로지스틱 회귀 분석과의 보다 정확한 비교를 위해 동일한 변수를 입력 데이터 값으로 하는 신경망 모형(모형 2)도 구축하여 세 모형의 예측율을 비교하였다. 그 결과, 신경망 모형 1이 가장 높은 분류적중율을 나타냈으며, 이 모형에 따라 성별, 컴퓨터사용시간, 가구월평균소득, 친한친구수, 비행친구수, 개인공부시간, 자기통제력, 사교육시간, 여가시간, 자기신뢰감, 스트레스, 학교적응, 공부고민 등의 변수들로 청소년의 컴퓨터 오락추구 행동을 예측하는 것이 보다 정확하고 효율적임을 제시하였다. 본 연구의 결과는 청소년의 컴퓨터 오락추구 행동을 예측하고 진단하거나 적절하게 조절·대처하는데 사용될 수 있음을 제안한다.

주제어 : 신경망, 컴퓨터 오락추구 행동, 데이터마이닝

Application of the Neural Network to Predict the Adolescents' Computer Entertainment Behavior

Hyejoo Lee[†] · Euihyun Jung^{††}

ABSTRACT

This study investigates the predictive model of the adolescents' computer entertainment behavior using neural network with the KYPS data (3449 in the junior high school; 1725 boys and 1724 girls). This study compares the results of neural network(model 1) to the logistic regression model and neural network(model 2) with the exact same variables used in logistic regression. The results reveal that the prediction of neural network model 1 is the highest among three models and with gender, computer use time, family income, the number of close friends, the number of misdeed friends, individual study time, self-control, private education time, leisure time, self-belief, stress, adaptation to school, and study related worries, the neural network model 1 predicts the computer entertainment behavior more efficiently. These results suggest that the neural network could be used for diagnosing and adjusting the adolescents' computer entertainment behavior.

Keywords : Neural Network, Computer Entertainment Behavior, Data Mining

[†] 정 회 원: 중앙대학교 교육학과 강사
^{††} 중신회원: 안양대학교 컴퓨터학과 부교수(교신지자)
논문접수: 2012년 09월 13일, 심사완료: 2013년 03월 01일, 게재확정: 2013년 03월 11일

1. 서론

현대사회에서 컴퓨터는 생활전반에 두루 활용되어 없어서는 안 되며, 사회전반에 큰 영향력을 미치는 중요한 존재로 자리매김하였는데, 이를 통해 인간의 삶에 있어 긍정적인 효과를 제공하기도 하고 동시에 다양한 부작용과 문제를 유발하기도 한다. 컴퓨터를 사용하는 유형 중 오락추구 행동도 이와 마찬가지로 스트레스 해소, 자신감 증진 등 인성발달 및 인지발달 등의 유익한 영향을 주기도 하지만, 인터넷 중독, 게임 중독이라는 부작용을 유발하기도 한다. 청소년은 보통 컴퓨터 오락에 대해 왕성한 선호도를 보이는데 이는 새로운 것에 대한 강한 매력을 느끼고 자신의 욕구와 스트레스를 분출하는 수단이자 놀이로서 생각하고 있기 때문이다. 이러한 이유로 청소년들에게 컴퓨터 오락이 미치는 영향은 매우 크다고 할 수 있으며, 적절한 오락추구 행동으로 그치지 않고 과도하게 몰입하게 되어 결국 중독되는 등 청소년이 가장 많은 문제를 일으키는 영역으로 지적되어 왔다[1][2][3]. 따라서 컴퓨터의 영향과 관련하여 오락추구 행동에 대해 관심을 기울여야 하며, 컴퓨터 오락추구 행동에 미치거나 예측하는 여러 변수들에 대한 폭넓은 이해가 필요하다.

청소년의 컴퓨터 오락추구 행동과 관련된 요인에 대해 연구마다 차이가 있지만 대체적으로 성별, 자기통제력, 자아존중감, 공격성, 생활만족도, 컴퓨터사용시간 등과 같은 개인관련 변수, 부모공부기대, 부모와의 관계, 형제관계, 가족구성, 경제적 상태, 부모학력 등과 같은 가정관련 변수, 그리고 학교생활 적응, 교사 및 또래와의 관계 등과 같은 학교관련 변수 등 다양한 측면이 관련되어 있다고 보고되었다[3][4][5][6]. 그러나 지금까지의 컴퓨터 오락추구 행동과 관련된 요인에 대한 연구들이 진행되어왔지만 아직 미흡한 편이며 컴퓨터 게임이나 인터넷 중독 등 부작용에 더 많이 집중하여, 일반적인 컴퓨터 오락추구 행동과 관련된 변수들을 분석한 연구는 드물다[7]. 또한 관련 변수들을 분석했다라도 연구자의 관심에 따라 특정 변수들을 선택하여 투입함으로써 상반된 결과들을 제시하고 있다. 즉, 기존 연구들은 다양한 변수들을 종합적으로 고려하는데 있어서 한계가

있었다.

데이터마이닝(Data Mining)은 대용량의 데이터에 내재되어 있는 유의미한 정보를 찾아내기 위해 자동적이거나 반자동적으로 데이터를 탐색, 분석하는 방법이다[6]. 데이터마이닝 기법 중 신경망(Neural Networks)은 복잡한 구조를 가진 데이터의 분류 및 예측에 사용되는 비선형모형으로써, 대용량의 복잡한 구조를 가진 데이터간 발생할 수 있는 모든 상관유형을 추출해내어 예측 모형을 구축하는데 효율적이다[8][9]. 이에 본 연구에서는 신경망을 활용하여 청소년의 컴퓨터 오락추구 행동을 설명하는 예측모형을 구축하고 예측의 목적으로 사용되었던 전통적인 통계방법인 로지스틱 회귀분석과의 비교를 통해 신경망의 성능을 분석하고자 한다. 이를 통해 청소년의 컴퓨터 오락추구에 대한 깊이 있는 이해와 컴퓨터 오락추구 행동을 미리 진단하고 적절히 조절하거나 대처하는 데 유용한 기초자료를 제시할 수 있을 것으로 기대한다.

2. 이론적 배경

2.1 컴퓨터 오락추구 행동 및 관련 변수들

컴퓨터를 선택하고, 사용하는 목적은 다양한데, 대체적으로 오락추구, 관계추구, 정보추구 행동으로 구분될 수 있다[6]. 컴퓨터 오락추구 행동이란 주로 오락, 게임과 관련된 목적으로 인터넷을 포함한 컴퓨터를 사용하는 경우를 말한다[10]. 일반적으로 오락이나 게임은 청소년들에게 있어 적절한 인성발달과 인지발달, 스트레스 해소 등에 도움을 주는 중요한 도구의 의미를 지니고 있다. 이러한 특성에 더하여, 컴퓨터 오락추구 행동은 컴퓨터의 보편화, 게임 프로그램 자체의 발전, 대안놀이 등의 부재로 청소년들의 일상생활이 되어버렸으며, 청소년문화를 대표하는 것으로 인식되고 있다[4]. 그러나 컴퓨터 오락추구 행동이 과도하게 되면, 청소년의 호기심과 승부욕을 자극하며 통쾌감을 주어 더욱 게임에 몰입하게 하고 중국에는 중독에 이르게 되어 청소년이 가장 많은 문제를 일으키는 영역으로 지적받아왔다[1][2][3]. 따라서 컴퓨터 오락추구 행동과 이에 영향을 미치는 변

수들에 더욱 관심을 기울일 필요성이 강조된다[2].

컴퓨터 오락추구 행동과 관련된 변수로, 개인, 가정, 학교 등 여러 측면이 고려되었으며, 연구마다 다소 차이를 나타냈다. 개인관련 요인으로는 성별, 자기통제력, 자아존중감, 공격성, 생활만족도, 컴퓨터사용시간 등이 연구되었으며, 그 결과, 대체적으로 남학생이 여학생보다 컴퓨터 오락추구 행동을 더 많이 나타냈으며 자기통제력과 자아존중감이 낮을수록, 스트레스가 높을수록, 공격성이 높을수록, 자신의 삶에 더 많이 만족할수록, 컴퓨터사용시간이 많을수록 컴퓨터 오락추구 행동에 더 쉽게 빠지는 것으로 보고되었다[3][5][6]. 가정과 관련된 특성은 심리적, 경제적으로 부모에게 의존하는 시기인 청소년에게 매우 중요한 요인으로 여겨지고 있다[4]. 가정관련 요인으로는 부모공부기대, 부모와의 관계, 형제관계, 가정의 구성이나 경제적 상태, 부모학력 등이 연구되었다. 그 결과 부모학력의 경우에는 상반된 결과가 있었지만, 대체로 부모공부기대가 높을수록, 부모와의 관계와 형제관계가 좋지 않을수록, 사회경제적 지위가 낮거나 경제적으로 취약한 가정일수록 컴퓨터 오락추구 행동을 증가하는 것으로 나타났다[4][5][6]. 학교관련 요인으로는 학교생활적응정도, 교사 및 또래와의 관계 등에 대해 연구되었으며, 그 결과, 학교생활에 대한 적응 정도나 교사와 또래의 지지나, 친한 친구 등과 컴퓨터 오락추구 행동간의 관련성에 대해 상반된 결과가 제시되었다[6]. 지금까지 컴퓨터 오락추구 행동과 관련된 요인에 대한 연구들이 진행되어왔지만 아직 미흡한 편이며[7], 관련 변수들간의 상호작용과 조합관계를 조사한 연구도 있었지만[6], 대부분의 연구에서는 각기 연구마다 관심 있는 특정 변수들을 중심으로 연구함으로써, 연구마다 일관되지 않은 결과를 제시하고 있어서 보다 정확하게 컴퓨터 오락추구 행동을 예측하기 위해서는 여러 변수들을 종합적으로 고려한 연구가 필요하다.

2.2 관련변수 추출 알고리즘을 위한 신경망

신경망은 복잡한 구조를 가진 데이터간에 발생할 수 있는 모든 상관유형을 함수에 포함시키는 병렬형 방식으로 인간두뇌의 학습과정을 시냅스

작용을 통한 연산과정으로 재현한 데이터마이닝 방법의 일종이다. 전형적인 신경망은 입력층, 출력층, 입력층과 출력층 사이의 은닉층으로 구성되어 있으며, 입력층에서는 신호를 받고, 출력층에서는 출력신호를 보내며, 은닉층은 입력층과 출력층을 연결시킨다. 이 방법은 실제 출력값과 기대 출력값을 비교하고, 그 차이를 시냅스 역할을 하는 노드(node)에 가중치를 주어 조절한 후, 그 과정을 신경망 구조가 안정화될 때까지 반복함으로써 예측모형을 자가 창출한다[8][9].

신경망은 데이터를 처리하는 은닉층이 블랙박스적 특성을 갖고 변수들간의 상관관계를 뚜렷이 표출하지 못하여 구성 노드들간의 관계를 해석하기 힘들다는 단점이 있음에도 불구하고 다양한 분야에서 예측 데이터를 얻기 위해서 유용하게 활용되고 있는데, 이는 신경망이 복잡한 비선형적인 관계를 표현할 수 있기 때문이다. 즉, 통계적 모형의 전제가 되는 여러 가지 가정을 충족하지 못하는 상황에서도 그 적용의 타당성이 보장될 수 있기 때문이다. 특히 많은 입력변수를 가지고 있으며, 입력변수와 출력변수 간의 관계가 복잡한 비선형 형태일 때 효율적이다[11][12].

3. 연구 방법

3.1 분석자료 및 분석대상

본 연구는 한국청소년정책연구원의 청소년패널 데이터(Korea Youth Panel Survey: KYPS) 중 2003년에 조사된 중 2 패널 1차 자료를 사용하여 분석하였다. KYPS는 전국의 중학교 2학년에 재학 중인 학생들을 표집 하여, 표집된 학생과 그들의 학부모를 대상으로 1년 간격으로 실시한 설문조사 결과이다. 표집에는 전국 초등학교 수 대비 지역별 중학교 학생 수를 층화하여 할당한 후, 지역별로 학교 수를 결정하고 추출한 뒤 1개 학급을 무작위로 추출하는 층화다단계군집표집(stratified multi-stage cluster sampling)이 이용되었다. 1차년도의 최종 표본은 총 3449명이었으며, 성별은 남학생 1725명, 여학생 1724명으로 1:1의 비율을 보였고, 거주 지역은 서울지역이 17%(585명), 서울 제외 도시지역이 76%(2622명),

농촌지역이 7%(242명)이었다.

3.2 조사도구

신경망 입력데이터를 결정하기 위하여 관련된 기존 연구들을 조사한 결과, 컴퓨터 오락추구 활동에 영향을 미치는 변수들은 크게 개인, 가정, 학교와 관련된 것으로 확인되었다. 데이터마이닝은 대용량의 데이터 중에서 예측 모형을 추출해 내고자 하므로 많은 수의 변수들을 예측변수로 다루게 된다. 이처럼 변수들의 수가 많을 때 후보 변수들간 정보가 상당히 중첩될 가능성이 있으므로 중복성이 크게 나타나는 변수들을 찾아내 미리 제거하였다[13]. 구체적인 신경망의 입력데이터는 <표 1>과 같다.

<표 1> 신경망의 입력데이터

영역	세부 입력데이터
개인관련(16)	성별, 자기통제력, 공격성, 자아존중감, 스트레스, 낙관적 성향, 자기인식문제아, 생활만족도, 건강상태, 심리정신상태, 사이버동아리 참여, 본인희망교육, 주변비난격정, 컴퓨터사용시간, 여가시간, 자기신뢰감
가정관련(9)	부모공부기대, 부모애착, 부모폭력, 부부간폭력, 형제관계, 집안경제상태, 부학력, 가족구성형태, 가구월평균소득
학교관련(10)	교사애착, 학교적응, 성적, 공부고민, 또래애착, 친구만남정도, 친한친구수, 비행친구수, 사교육시간, 개인공부시간

신경망의 입력데이터는 그대로 이용하지 않고 분석에 적합한 형식으로 변환하여 이용하는데, 데이터가 0과 1 사이의 값을 가질 때 수렴 속도가 빠르고 학습도 원활하다[11]. 숫자정보인 데이터는 연속형과 범주형으로 나눌 수 있는데, 연속형은 표준화하여 입력 변수값을 0과 1 사이에 있도록 조정하며, 표준화 수식은 식(1)과 같다. 본 연구에서 연속형 데이터는 식(1)에 따라 표준화하였다(예: 컴퓨터사용시간, 개인공부시간 등).

$$\text{표준화된 변수값} = \frac{\text{실제값} - \text{최소값}}{\text{최대값} - \text{최소값}} \quad (1)$$

범주형 데이터 중 순서적 의미를 가지는 데이터는 0과 1 사이의 값으로 범주화하여 지정하였고, 순서적 의미 없는 데이터는 더미(dummy) 변

수로 변환하여 해당되는 변수면 1, 아니면 0으로 지정하였다(예: 성별, 사이버동아리 참여 등)[11].

본 연구에서 사용된 종속변수 또는 출력변수는 컴퓨터 오락추구 행동에 대하여 자주 하는 편이다(1) 또는 그렇지 않다(0)로 변환하여 분석하였다.

3.3 분석방법

신경망을 위해 본 연구에서는 다층인식모형(MLP)을 사용하였다. 다층인식모형은 가장 널리 사용되고 있는 성공적인 응용아키텍처로서, 역전파(backpropagation) 알고리즘의 학습방법으로 입력층과 출력층 사이에 가상적인 은닉층을 설정하고 이를 통과하는 가중치를 반복적으로 수정하여 예측모형을 구축하게 된다. 이를 위해 데이터를 무작위로 훈련 데이터(train data)와 검정 데이터(test data)로 나눈 후, 훈련 데이터를 학습시켜 훈련 데이터의 레코드(record) 하나에 대해 학습시키고 그 오차를 신경망에 반영한다. 즉, 데이터의 한 레코드의 필드값들을 신경망에 입력시키고, 각 계층을 거쳐 계산된 출력값과 그 레코드의 목표값의 오차를 구한 후, 신경망의 출력층부터 은닉층을 거쳐 입력층에 이르기까지 역순으로 오차를 연결강도에 반영시킨다. 이와 같이 모든 훈련 데이터의 레코드에 대해 같은 작업을 반복적으로 수행한다. 이 과정을 통해 신경망의 연결강도가 결정되고 나면, 검정 데이터를 사용하여 구축한 신경망의 예측율을 파악한다[11][12]. 이러한 과정에 따라 본 연구에서는 데이터의 70%를 훈련 모형에 무작위로 투입하여 신경망 모형을 학습하였고, 그리고 나서 30%의 검정 집단에 이러한 모형에 대해 예측율(predictive accuracy)을 검증하였으며, 신경망 모형의 특성상 데이터를 분석할 때마다 무작위로 표본을 선정하기 때문에 다양한 모형을 생성해본 후, 최적으로 판단된 모형을 선택하였다.

또한, 본 연구에서는 컴퓨터 오락추구 행동을 예측하는데 있어서 신경망 활용의 효율성을 파악하기 위하여 두 개의 분석기법, 즉 신경망과 로지스틱 회귀분석을 비교하였다. 로지스틱 회귀분석은 비선형적인 관계를 보이는 현상들의 예측과

분류 문제에 자주 사용되는 통계적 분석기법으로, 이러한 방식은 특정 행동을 예측하는데 있어서 신경망 효율성을 분석하기 위해 여러 연구들에서 활용되었다[8][11][12].

신경망은 비선형적 패턴들을 찾도록 설계되었으며, 결측치를 다루는데 매우 적합하기 때문에 신경망 모형에서는 연속형과 범주형 입력데이터 모두를 활용하여 분석가능하다. 따라서 신경망 분석에는 <표 1>의 입력데이터가 포함되었다(총 35개). 추가적으로 본 연구에서는 예측모형을 구축하는데 있어서 신경망과 로지스틱 회귀분석의 성능을 비교하고자 하였으며, 두 방법 간의 정확한 비교를 위하여 본래의 신경망 분석 외에 로지스틱 회귀분석에서 사용되는 변수를 같은 입력데이터 값으로 하는 신경망 분석도 수행될 필요가 있었다[8]. 따라서 본 연구에서는 로지스틱 회귀분석을 하기 위해 신경망 분석에 포함된 입력데이터 중 연속형 범주를 구간으로 나눠서 더미변수로 변환하였고, 그 결과 범주형 변수를 포함하여 총 63개의 입력데이터가 로지스틱 회귀분석에 포함되었고, 예측모형에 대한 신경망 성능의 분석을

위해 연속형과 범주형 변수를 입력데이터 값으로 사용하는 일반적인 신경망(모형 1)과 로지스틱 회귀분석에서 사용된 더미변수를 똑같이 입력데이터 값으로 활용한 신경망(모형 2)을 수행하였다.

일반적으로 로지스틱 회귀분석에서 다중공선성이 존재하는 경우, 서로 높은 상관을 보이는 변수들이 종속변수에 미치는 영향을 정확히 파악할 수 없다[14]. 이에 본 연구에서는 회귀모형에 포함된 어떠한 독립변수들도 다중공선성의 문제를 야기하지 않음을 확인하였다. 신경망 모형들은 SPSS의 Clementine 6.0, 로지스틱 회귀분석은 SPSS 12.0을 사용하여 측정하였다.

4. 분석 결과

다층인식모형(MLP) 신경망을 적용시키기에 앞서 은닉층(hidden layer)과 은닉층 노드의 개수를 정해야 한다. 은닉층과 은닉층 노드수가 많아지면 모형이 복잡해지고 추정해야 할 계수의 수가 급격히 증가하여 최적화가 어려워진다. 은닉층과 은닉층 노드수의 결정은 다양한 값을 시도한 후에

<표 2> 신경망 모형 1 결과

	예측율(%)		전체 예측율(CCR)
	오락추구형	기타유형	
오락추구형	90.4	9.6	89.0
기타유형	12.6	87.4	
모형의 구조			
입력층	35 노드		
은닉층 #1	9 노드		
은닉층 #2	8 노드		
출력층	1 노드		
Input	중요도(정규화 중요도)	Input	중요도(정규화 중요도)
성별	0.107(100.0%)	교사애착	0.022(20.4%)
컴퓨터사용시간	0.058(54.4%)	친구만남정도	0.022(21.0%)
가구월평균소득	0.046(43.4%)	성적	0.021(19.3%)
친한친구수	0.046(43.3%)	자아존중감	0.021(19.2%)
비행친구수	0.045(42.2%)	부모애착	0.019(17.5%)
개인공부시간	0.045(42.0%)	자기인식문제아	0.019(17.5%)
자기통제력	0.044(40.8%)	건강상태	0.018(17.1%)
사교육시간	0.044(40.8%)	부모학대	0.018(17.1%)
여가시간	0.040(37.1%)	부학력	0.018(16.6%)
자기신뢰감	0.032(29.6%)	공격성	0.017(15.7%)
스트레스	0.030(28.3%)	주변비난걱정	0.017(15.4%)
학교적응	0.030(28.4%)	부부간학대	0.016(14.6%)
공부고민	0.029(27.5%)	또래애착	0.013(12.6%)
생활만족도	0.023(21.9%)	가족구성형태	0.013(12.4%)
부모공부기대	0.023(21.9%)	낙관적 성향	0.013(12.1%)
집안경제상태	0.023(21.3%)	본인희망교육	0.012(10.7%)
심리정신상태	0.022(21.0%)	사이버동아리 참여	0.011(10.3%)
형제관계	0.022(20.3%)		

가장 최적을 판단되는 값으로 결정하고 그 값을 검증하는데, MLP 모형은 전형적으로 하나 또는 두 개의 은닉층을 갖는다[15]. 본 연구에서는 신경망의 수치연산에 있어서 더 적합한 구조로 알려져 있으며, 오차율이 더 적은 경향을 보인다는 기존연구를 토대로[16][17] 은닉층을 2층으로 결정하였다. 또한 은닉층의 노드수가 많아지면 그 만큼 연산량도 증가하므로 각 신경망의 은닉층 노드수를 1개부터 입력데이터의 개수만큼 노드수를 증가시켜 실험하여, 예측율이 높고 오류가 가장 적은 은닉층의 노드 개수를 결정하였다.

신경망 학습은 검증결과와 훈련결과의 시스템

오차(EP)가 0.05일 때의 학습회수에 근거하여 결정하였다. 신경망 모형 1의 학습회수는 15790이었으며, 그 결과는 <표 2>에 제시하였다. 신경망 모형 1의 입력층은 각 입력변수에 대응되는 35개의 노드로, 은닉층 1층은 9개의 노드로, 은닉층 2층은 8개의 노드로 구성된 구조를 나타냈다. 신경망 모형 1은 컴퓨터 오락추구형 학생들의 90.4%를 정확하게 분류하였고, 기타유형의 87.4%를 정확하게 분류하였다. 또한 훈련된 신경망 모형 1로 컴퓨터 오락추구 행동을 예측하는데 사용되는 경우, 전체 예측율(overall correct classification rate, CCR)은 89.0%를 나타냈다. 입력데이터의

<표 3> 신경망 모형 2 결과

	예측율(%)		전체 예측율(CCR)
	오락추구형	기타유형	
오락추구형	84.4	15.6	83.6
기타유형	17.4	82.6	
모형의 구조			
입력층		63 노드	
은닉층 #1		9 노드	
은닉층 #2		8 노드	
출력층		1 노드	
Input	중요도(정규화 중요도)	Input	중요도(정규화 중요도)
성별	0.108(100.0%)	사이버동아리 참여	0.013(12.5%)
비행친구수 20명이상	0.032(29.5%)	낙관적 성향	0.013(12.5%)
여가시간 10시간이상	0.030(27.9%)	자기인식문제아	0.013(12.0%)
스트레스	0.026(23.9%)	비행친구수 5명미만	0.013(11.9%)
교사애착	0.024(22.4%)	자기통제력	0.013(11.7%)
부모공부기대	0.022(20.5%)	공격성	0.012(11.5%)
개인공부시간 8~10시간미만	0.022(20.5%)	가구월평균소득 400~600만원이하	0.012(11.5%)
가족구성형태	0.022(20.3%)	가구월평균소득 200~400만원미만	0.012(11.3%)
자아존중감	0.021(19.4%)	가구월평균소득 200만원미만	0.012(11.1%)
성적	0.020(18.1%)	자기신뢰감	0.012(10.7%)
컴퓨터사용시간 6시간이상	0.019(17.2%)	사교육시간 12~16시간미만	0.012(10.7%)
형제관계	0.018(16.4%)	여가시간 8~10시간미만	0.011(10.4%)
컴퓨터사용시간 4~6시간미만	0.018(16.4%)	개인공부시간 12시간이상	0.011(10.4%)
부모학대	0.017(16.0%)	친한친구수 20명이상	0.011(10.4%)
컴퓨터사용시간 2시간미만	0.017(15.5%)	주변비난걱정	0.011(10.2%)
부학력	0.017(15.5%)	공부고민	0.011(10.2%)
친구만남정도	0.017(15.3%)	개인공부시간 4~6시간미만	0.011(10.2%)
개인공부시간 2시간미만	0.016(15.0%)	가구월평균소득 600만원이상	0.011(9.9%)
여가시간 6~8시간미만	0.016(15.3%)	사교육시간 20시간이상	0.011(9.9%)
개인공부시간 6~8시간미만	0.016(14.8%)	비행친구수 5~10명미만	0.010(9.7%)
여가시간 2시간미만	0.016(14.6%)	사교육시간 8~12시간미만	0.010(9.3%)
개인공부시간 10~12시간미만	0.015(14.0%)	비행친구수 10~15명미만	0.009(8.6%)
학교적응	0.015(14.0%)	심리정신상태	0.009(8.6%)
개인공부시간 2~4시간미만	0.015(13.6%)	건강상태	0.009(8.4%)
사교육시간 4시간미만	0.014(13.4%)	본인희망교육	0.009(8.4%)
여가시간 2시간~4시간미만	0.014(13.3%)	사교육시간 16~20시간미만	0.009(8.5%)
사교육시간 4시간~8시간미만	0.014(13.1%)	컴퓨터사용시간 2~4시간미만	0.009(8.3%)
부모애착	0.014(12.8%)	생활만족도	0.008(7.8%)
여가시간 4시간~6시간미만	0.014(12.7%)	부부간학대	0.008(7.2%)
집안경제상태	0.014(12.7%)	또래애착	0.007(6.7%)
비행친구수 15~20명이하	0.014(12.7%)	친한친구수 10~20명미만	0.007(6.5%)
친한친구수 10명미만	0.014(12.6%)		

<표 4> 로지스틱 회귀 모형 결과

변수	예측율(%)		전체 예측율(CCR)					
	오락추구형	기타유형	B	S.E	Wald	자유도	유의확률	Exp(B)
오락추구형	79.7	20.3					79.8	
기타유형	20.1	79.9						
자기통제력								
공격성			-.448	.154	8.448	1	.004	.639
자아존중감			.155	.260	.358	1	.550	1.168
스트레스			-.252	.188	1.788	1	.181	.777
낙관적 성향			.819	.289	8.012	1	.005	2.268
자기인식문제해			.088	.144	.374	1	.541	1.092
생활만족도			-.375	.457	.674	1	.412	.687
건강상태			.054	.144	.141	1	.708	1.056
심리정신상태			.105	.143	.532	1	.466	1.110
사이버동아리 참여			-.048	.189	.064	1	.801	.953
본인희망교육			-.148	.154	.930	1	.335	.862
주변비난걱정			.036	.370	.010	1	.922	1.037
컴퓨터사용시간 2시간미만			-.123	.143	.740	1	.390	.884
컴퓨터사용시간 2~4시간미만			-.581	.384	2.296	1	.130	.559
컴퓨터사용시간 4~6시간미만			-.111	.371	.089	1	.766	.895
여가시간 2시간미만			.150	.399	.141	1	.707	1.162
여가시간 2~4시간미만			-.1030	.448	5.280	1	.022	.357
여가시간 4~6시간미만			-.1207	.409	8.717	1	.003	.299
여가시간 6~8시간미만			-.828	.404	4.196	1	.041	.437
여가시간 8~10시간미만			-.623	.421	2.193	1	.139	.536
자기신뢰감			-.929	.457	4.134	1	.042	.395
부모공부기대			.061	.145	.177	1	.674	1.063
부모애착			.381	.135	7.953	1	.005	1.463
부모학대			.146	.146	1.002	1	.317	1.158
부부간학대			-.615	.391	2.477	1	.116	.541
형제관계			-.379	.343	1.222	1	.269	.685
집안경제상태			.023	.135	.028	1	.866	1.023
부학력			-.205	.171	1.430	1	.232	.815
가족구성형태			.045	.140	.104	1	.747	1.046
가구월평균소득 200만원미만			-.389	.347	1.259	1	.262	.677
가구월평균소득 200~400만원미만			-.111	.308	.129	1	.720	.895
가구월평균소득 400~600만원미만			.069	.266	.068	1	.795	1.072
또래애착			-.115	.284	.163	1	.686	.892
교사애착			-.096	.156	.380	1	.537	.908
공부고민			-.182	.300	.369	1	.543	.834
학교적응			-.287	.133	4.640	1	.031	.750
친구만남경도			.371	.171	4.692	1	.030	1.450
친한친구수 10명미만			-.165	.199	.686	1	.408	.848
친한친구수 10~20명미만			-.466	.245	3.620	1	.057	.628
비행친구수 5명미만			-.186	.260	.512	1	.474	.830
비행친구수 5~10명미만			1.174	.351	11.159	1	.001	3.235
비행친구수 10~15명미만			1.012	.398	6.468	1	.011	2.750
비행친구수 15~20명미만			.705	.456	2.387	1	.007	2.023
사교육시간 4시간미만			1.422	.529	7.217	1	.069	4.144
사교육시간 4~8시간미만			-.646	.355	3.315	1	.069	.524
사교육시간 8~12시간미만			-.685	.332	4.241	1	.039	.504
사교육시간 12~16시간미만			-.493	.289	2.919	1	.088	.611
사교육시간 16~20시간미만			-.264	.285	.857	1	.355	.768
개인공부시간 2시간미만			-.206	.345	.356	1	.551	.814
개인공부시간 2~4시간미만			.267	.264	1.028	1	.311	1.307
개인공부시간 4~6시간미만			-.099	.248	.160	1	.689	.906
개인공부시간 6~8시간미만			.038	.257	.022	1	.881	1.039
개인공부시간 8~10시간미만			-.017	.268	.004	1	.948	.983
개인공부시간 10~12시간미만			-.297	.297	.999	1	.317	.743
성적			.177	.377	.219	1	.640	1.193
성별			.315	.154	4.210	1	.040	1.370
(Constant)			2.811	.138	415.953	1	.000	16.618
			-.080	.953	.007	1	.933	.923

-2LL=1604.515
 $\chi^2=728.113(df=56, p=.000)$, Nagelkerke $R^2=.47$

상대적 중요도(정규화 중요도)를 보면, 성별, 컴퓨터사용시간, 가구월평균소득, 친한친구수, 비행친구수, 개인공부시간, 자기통제력, 사교육시간, 여가시간, 자기신뢰감, 스트레스, 학교적응, 공부고민 등이 높은 영향력을 미치는 순서로 결과가 나타났다.

로지스틱 회귀 모형에서와 같은 변수를 사용한 신경망 모형 2의 학습회수는 5882이었으며, 그 결과는 <표 3>과 같다. 훈련된 모형의 전체 예측율은 83.6%로 신경망 모형 1보다는 낮았다. 신경망 모형 2는 컴퓨터 오락추구 행동을 나타내는 학생들의 84.4%를 정확하게 분류하였고, 기타유형은 82.6%의 분류 예측율을 보였다. 신경망 모형 2의 입력층은 각 입력변수에 대응되는 63개의 노드로, 은닉층 1층은 9개의 노드로, 은닉층 2층은 8개의 노드로 구성된 구조를 나타냈다. 입력데이터의 상대적 중요도(정규화 중요도)를 보면 영향력의 면에서 성별, 비행친구수 20명이상, 여가시간 10시간이상, 스트레스, 교사에착 등의 순으로 나타났다.

신경망 모형과의 비교를 하기 위하여 로지스틱 회귀분석을 실시하였으며, 그 결과는 <표 4>에 제시하였다. 실제 관측치와 예측치를 비교 분석하여 로지스틱 모형의 적합성을 살펴보는 Hosmer-Lemeshow 적합도($\chi^2=7.96$, $df=8$, $p=.44$)를 살펴본 결과, 실제 관측치와 모형에 의한 기대치의 차이가 거의 없는 것으로 나타나 본 분석모형이 적합함을 알 수 있었다. 모형의 적절성을 나타내주는 Cox와 Snell R^2 은 .35이고 Nagelkerke R^2 은 .47이었다. 로지스틱 회귀분석 결과, 본 분석에 투입된 변수로 청소년의 오락추구 행동을 예측할 확률은 79.7%이었고, 기타유형을 예측할 확률은 79.9%이었으며, 전체 예측율(CCR)은 79.8%로 나타나, 전체 예측율(CCR)에서 신경망 모형 1과 2보다 더 낮게 제시되었다.

<표 4>를 보면, 청소년의 컴퓨터의 오락추구 행동을 예측하는데, 총 15개의 변수가 유의한 영향을 미치는 것으로 나타났는데, 승산비(Exp(B))에 비추어볼 때, 컴퓨터의 오락추구 행동에 가장 큰 영향을 미치는 변수는 성별이었으며, 비행친구수 5명미만, 비행친구수 5~10명미만, 스트레스,

비행친구수 10~15명미만, 부모공부기대, 학교적응, 성적, 공부고민, 자기통제력, 여가시간 4~6시간미만, 여가시간 8시간이상, 여가시간 2시간미만, 여가시간 2~4시간미만의 순으로 영향을 미치는 것으로 나타났다. 이 중에서 성별, 비행친구수, 스트레스, 부모공부기대, 학교적응, 성적은 컴퓨터 오락추구 행동을 할 승산비 증가에 유의하게 기여하는 반면, 이외 여가시간, 자기통제력, 공부고민은 승산비 감소에 유의하게 기여하는 것으로 나타났다.

5. 결론

본 연구에서는 신경망을 청소년의 컴퓨터 오락추구 행동에 대한 예측모형을 구축하는데 사용하였다. 본 연구에서는 컴퓨터 오락추구 행동에 미치는 개인, 가정, 학교관련 변수들을 추출하여 입력데이터로 포함한 신경망을 구축하고(모형 1), 이를 전통적인 통계방법의 하나로 예측에 유용한 로지스틱 회귀분석과의 비교를 통해 신경망의 성능을 분석하였다. 보다 정확한 비교를 위해 연속형과 범주형 입력데이터를 모두 활용한 신경망 모형 외에(모형 2) 로지스틱 회귀분석과 동일한 변수를 입력데이터 값으로 하는 신경망 모형 2도 함께 구축하였다.

신경망 모형 1과 모형 2, 그리고 로지스틱 회귀모형에 따라 청소년의 컴퓨터 오락추구 행동에 대해 상대적으로 중요한 변수들이 추출되었는데, 각 변수들은 모형에 따라 조금씩 다른 것으로 나타났다. 그러나 신경망 모형 1이 세 모형 중 89.0%의 가장 높은 분류적중율을 나타내었으므로, 신경망 모형 1에서 추출된 변수들로 청소년의 컴퓨터 오락추구 행동을 예측하는 것이 보다 더 정확하다고 할 수 있겠다. 신경망 모형 1에서 추출된 청소년의 오락추구 행동에 상대적으로 중요한 영향을 미치는 변수들은 성별, 컴퓨터사용시간, 가구월평균소득, 친한친구수, 비행친구수, 개인공부시간, 자기통제력, 사교육시간, 여가시간, 자기신뢰감, 스트레스, 학교적응, 공부고민 등이었다. 이러한 결과는 대상 행동을 예측하는데 있어서 신경망 모형이 로지스틱 회귀 모형보다 더 우수했다고 보고한 연구들과 일치한다[12].

따라서 본 연구에서 구축한 예측모형을 활용하여 컴퓨터 오락추구 행동을 보다 정확하게 이해할 수 있을 뿐만 아니라, 오락추구 행동을 예측하고, 이에 집착하거나 중독되지 않도록 미리 진단하고 적절히 조절하거나 대처하는데 사용될 수가 있다. 이를 통해 교육자나 부모, 교육행정가 등이 학습자의 오락추구 행동을 방지하거나 감소, 또는 조절해야 할 것인지를 결정할 수가 있다. 또한 학습자의 오락추구 행동이 오락중독으로 진행되지 않도록 보다 효과적이고 효율적으로 조절, 중재하는 전략이나 프로그램 등에 본 연구의 결과를 활용할 수 있을 것이다[8]. 아울러 전통적인 분석방법에 비해 신경망 모형이 측정의 제약이나 방법적인 한계를 상당부분 극복할 수 있으며, 학습자의 특정 행동을 예측하는데 유용한 수단이 될 수 있다는 것을 보여주고 있어 앞으로 교육 분야에 신경망 모형의 적용을 활성화시키는데 기여할 것을 기대한다.

본 연구의 제한점과 추후연구에 대해 제언하면 다음과 같다. 본 연구에서는 상대적으로 신경망 모형의 설명력이 우수하였음을 밝혔으며, 이는 신경망이 다양한 분야에서 예측 데이터를 얻기 위하여 유용하게 활용됨을 제시하였다. 하지만, 신경망 모형은 각 구성 노드들 간의 관계를 해석하기 힘들다는 단점이 있다[12]. 따라서 후속 연구에서는 보다 향상된 신경망 구조로부터 규칙을 추출하여 컴퓨터 오락추구 행동에 영향을 미치는 중요한 요소를 보다 쉽게 관별하는데 초점을 둘 필요가 있으며, 또한 회귀모형 등 기존의 통계방법이나 의사결정트리 등 다른 데이터마이닝의 방법을 통합하는 상호 보완적 방식으로 컴퓨터 오락추구 행동을 더 정확하게 예측하도록 노력해야 할 필요가 있다. 또한 본 연구에서는 전국 학생을 대상으로 표집된 한국청소년패널조사의 데이터를 활용하여 분석을 하였다. 따라서 연구결과의 일반화 정도는 높다고 할 수 있다[14]. 그러나 본 연구의 측정도구가 모두 자기보고식 방법에 근거하고 있어서 컴퓨터 오락추구 행동과 관련 변수들을 객관적으로 측정하는데 한계가 있다. 추후 연구에서는 이를 보완한 보다 실제적인 측정이 요구된다.

참 고 문 헌

- [1] 정기원 · 권량희 (2006). 부모 및 교사 애착과 청소년 비행의 관계: 일반긴장이론의 실증분석. **한국청소년연구**, 13(5), 127-135.
- [2] 최나야 · 한유진 (2006). 충동성, 부모와의 의사소통 및 인터넷 게임에 대한 기대가 아동, 청소년의 게임 중독에 미치는 영향. **한국가정관리학회지**, 24(2), 209-219.
- [3] Young, K. S. (1996). Internet addiction: The emergence of a new clinical disorder. *Cyber Psychology and Behavior*, 1(3), 237-244.
- [4] 김기숙 · 김경희 (2010). 의사결정나무 분석기법을 이용한 청소년의 인터넷게임 중독 영향요인 예측 모형 구축. **대한간호학회지**, 40(3), 378-388.
- [5] 나은영 · 송종현 (2006). 어린이의 인터넷 · 컴퓨터게임 몰입에 미치는 자기통제성의 매개역할과 어머니의 개방적 커뮤니케이션의 영향. **한국언론학보**, 50(2), 116-147.
- [6] 이혜주 · 정의현 (2012). 의사결정트리 기반의 분석을 통한 청소년의 컴퓨터 사용 유형별 관련 변수 추출. **한국컴퓨터교육학회 논문지**, 15(2), 9-18.
- [7] 최명선 · 최태산 · 안재영 (2007). 가정의 사회적, 물리적, 심리적 환경 변인과 청소년 인터넷 게임중독 성향간의 관계. **한국가족복지학**, 12(2), 5-27.
- [8] González, J., & DesJardins, S. (2002). Artificial neural networks: A new approach for predicting application behavior. *Research in Higher Education*, 43(2), 235-258.
- [9] Vandamme, J. P., Meskens, N., & Superby, J. F. (2007). Predicting academic performance by data mining methods. *Education Economics*, 15(4), 405-419.
- [10] 이선주 · 김귀애 · 홍창희 (2011). 인터넷 사용동기와 스트레스 대처방식이 청소년의 병리적 인터넷 사용에 미치는 영향: 성별차이를 중심으로. **한국심리학회지: 여성**, 16(3), 265-284.

- [11] 구자용 · 박헌진 · 최대우 · 김성수 (2001). **데이터마이닝**. 서울: 한국방송통신대학교출판부.
- [12] 오지영 · 이수정 (2008). 신경망을 이용한 초등학생 컴퓨터 활용 능력 예측. **정보교육학회논문지**, 12(3), 267-274.
- [13] Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. San Francisco, CA: Morgan Kaufman.
- [14] Allison, P. D. (1999). *Multiple regression: A primer*. Thousand Oaks, CA: Pine Forge Press.
- [15] Lippmann, R. P. (1988). An introduction to computing with neural nets. *Neural Networks*, 1, 371-377.
- [16] 윤길립 · 김병탁 (2002). 인공신경망 기법을 이용한 국내 해성점토의 압축특성 분석. **대한토목학회논문집**, 22(2), 133-147.
- [17] 홍원표 · 김원영 · 송영석 · 임석규 (2004). 인공신경망 모델을 이용한 산사태 예측. **한국지반공학학회논문집**, 20(8), 67-75.



이혜주

1992 동덕여자대학교 아동학과(이학사)
 1995 동덕여자대학교 아동학과(이학석사)
 2005 이화여자대학교 초등교육과(문학박사)
 2005 한국교육개발원 부연구위원대우
 2004~현재 중앙대학교 교육학과 강사
 관심분야: 창의성교육, u-러닝, 교수방법
 E-Mail: ladyzen@naver.com



정의현

1992 한양대학교 전자공학과(공학사)
 1994 한양대학교 전자공학과(공학석사)
 1999 한양대학교 전자공학과(공학박사)
 1999~2002 대우통신 선임연구원
 2002~2003 SCT 연구소장
 2004~현재 안양대학교 컴퓨터학과 부교수
 관심분야: IoT, 시맨틱웹, 데이터마이닝 등
 E-Mail: jung@anyang.ac.kr