

Group Mnet 기계학습 기법을 통한 중학생의 끈기(grit) 관련 변수 탐색*

유진은** · 김형관*** · 노민정****

초 록

청소년의 끈기(그릿)에 대한 기존 연구는 이론적으로 확인된 소수의 변수에 대하여 전통적인 분석 기법을 적용한 연구가 대부분이었다. 그러나 청소년의 끈기와 관련 있는 변수를 총체적인 관점에서 탐색하고 파악하여 새로운 변수를 발굴할 필요가 있다. 특히 한국아동·청소년패널조사(KCYPS) 2018은 끈기를 측정하는 문항을 새롭게 수집함으로써 청소년의 끈기 연구에 획기적인 계기를 제공하였다. 본 연구는 기계학습 기법을 통하여 KCYPS 2018의 수백 개의 변수를 한 모형에서 탐색함으로써 끈기와 관련 있는 변수를 총체적으로 파악하는 동시에 예측력이 높은 모형을 구축하고자 하였다. 기계학습 기법 중 벌점회귀모형은 예측력을 고려하면서도 상대적으로 해석하기 쉽고 설명하기 쉬운 모형을 도출한다. 벌점회귀모형 중 최신 기법인 Group Mnet은 LASSO의 회귀계수 불일치성 문제와 MCP의 다중공선성 문제를 다루며, 범주형 변수 또한 처리할 수 있다는 장점이 있다. 본 연구에서는 자료 분할로 야기되는 임의성을 통제하고자 100번의 반복을 통하여 선택계수(selection counts)를 구하고, 그 결과 301개의 설명변수 중 42개의 변수를 최종적으로 선택하였다. 선택된 변수들 중 선행연구에서 사용되었던 변수들은 학업 열의, 학업 무기력, 삶의 만족도, 행복감, 창의적 성격, 건강상태, 양육방식, 친구관계, 교사관계, 성별 등이었다. 선행연구에서는 찾아 볼 수 없었으나 선택된 변수로 자아존중감, 협동심, 우울, 주의집중, 공격성, 스마트폰 관련 변수, 팬덤 활동 빈도 등이 있었다. 본 연구의 의의를 기계학습 기법의 특징과 관련하여 논하고, 후속 연구 과제를 제시하였다.

주제어: 끈기, 그릿, 벌점회귀모형, 기계학습, Group Mnet, 한국아동·청소년패널(KCYPS 2018)

* 본 논문은 한국청소년정책연구원에서 수행한 한국아동·청소년패널조사(KCYPS 2018) 데이터를 활용하였으며, 제8회 한국아동·청소년 패널학술회의에 발표한 내용을 수정·보완한 것임.

** 한국고원대학교 제1대학 교육학과 교수, 교신저자, jeyoo@knue.ac.kr

*** 한국고원대학교 교육과정 및 교육평가 전공 석사

**** 한국고원대학교 박사, 원양초등학교 교사

I. 서 론

교육학 분야의 전통적인 연구 주제는 학업성취도와 같은 인지적인 특성이었는데, 정의적 특성과 인지적 특성의 연관성이 규명되기 시작하면서 정의적 특성에 대한 연구 또한 활발하게 이루어지고 있다. 그중 최근 대중에게도 널리 소개되며 관심이 집중되고 있는 정의적 특성이라면 단연 ‘grit’, 즉 끈기라 할 수 있을 것이다. Duckworth, Peterson, Matthews & Kelly(2007)가 처음 주창한 grit은 장기 목표에 대한 열정(passion)과 지속성(perseverance)으로 정의된다. 즉, 실패, 역경, 정체에도 불구하고 수년에 걸쳐 노력과 흥미를 지속하며 도전해 나가는 정의적 특성이 ‘grit’인 것이다.

성공 경험 또는 성취를 축적함으로써 삶의 목표와 의미를 찾아가야 하는 시기인 아동기와 청소년기에 끈기는 특히 더 중요하다. 타고난 재능이 있다고 하더라도 노력 없이는 성취를 이룰 수 없는데, 노력을 열정적으로 지속하려면 grit, 즉 끈기가 필요하기 때문이다(Duckworth, 2016). 끈기가 최근 소개된 개념이므로 다른 중요한 교육 관련 구인에 비해 상대적으로 연구가 많지 않았는데, 승인통계인 한국청소년연구원의 한국아동청소년패널조사(Korea Children and Youth Panel Study; 이하 KCYPS) 2018에서 끈기를 측정하기 시작함으로써 우리나라 아동청소년기 끈기 연구에 있어 획기적인 돌파구를 마련하였다.

수백 개의 문항으로 구성된 KCYPS 2018 원년 자료를 가능한 한 모두 활용하여 분석함으로써 우리나라 중학생의 끈기와 관련 있는 변수를 총체적인 관점에서 탐색하고 파악하는 것이 본 연구의 주된 연구목적이었다. 지금까지 사회과학 연구는 이론 또는 선행연구에 기반한 소수의 변수에 대하여 기존 통계기법을 활용하는 통계적 검정을 통하여 통계적 유의성을 확인하는 것이 관례였다. 그러나 기존 통계기법은 다음과 같은 문제가 있다. 첫째, 변수 조합에 따라, 심지어 어떤 변수를 먼저 투입하느냐에 따라 통계적 유의성이 달라질 수 있다. 둘째, 기존 통계기법을 활용하여 수백 개의 변수를 한 모형에 투입하려 할 경우 비수렴(nonconvergence) 및 과적합(overfitting) 등의 문제가 발생 가능하다(Yoo, 2018). 셋째, 기존 기법은 ‘설명’에 치중하며 ‘예측’(prediction)은 도외시한다는 단점이 있다. 즉, 모형을 구축한 자료에 대한 설명력이 높은 반면, 새로운 자료에 대해서는 모형이 들어맞기 힘들다(유진은, 2019).

기계학습 기법을 활용할 경우 비수렴 또는 과적합 문제를 최소화하며 수백 개의

변수를 한 모형에서 한꺼번에 고려할 수 있다. 따라서 기존 통계기법에서 변수 조합 또는 변수 투입 순서를 달리 할 때 회귀계수의 부호가 바뀌는 것과 같은 문제를 줄일 수 있다. 또한, 새로운 자료에도 잘 들어맞는, 즉 예측력을 고려하는 모형을 도출할 수 있다. 정리하자면, 본 연구는 기계학습 기법을 활용하여 KCYPS 2018 자료가 제공하는 수백 개의 변수를 한 모형에서 탐색함으로써 끈기와 관련 있는 변수를 총체적으로 탐색하여 새로운 변수를 파악하는 동시에 예측력이 높은 모형을 구축하고자 하였다. 기계학습 기법은 예측을 중시하므로 연구 결과의 일반화 또한 용이하게 된다는 장점이 있다(유진은, 2019; Yarkoni & Westfall, 2017). 본 연구에서는 기계학습 기법 중 벌점회귀모형(penalized regression)으로 분류되는 최신 기법인 Group Mnet (Huang, Breheny, Lee, Ma & Zhang, 2016)을 활용하였다.

II. 선행연구

본 연구의 반응변수인 KCYPS 2018의 끈기(grit)는 2018년 처음 수집된 자료를 2019년에 공개한 것이므로 2019년 12월을 기준으로 학술지에 출판된 KCYPS 2018 끈기 연구가 아직 없다. KCYPS 2018의 끈기는 김희명과 황매향(2015)이 Duckworth의 아동용 끈기 척도를 우리말로 타당화한 것이다. RISS와 Google scholar에서 김희명과 황매향(2015)의 끈기 척도를 활용한 네 편의 논문, 그리고 RISS에서 ‘grit’과 ‘중학생’ 또는 ‘고등학생’으로 검색하여 청소년의 grit을 연구한 논문 열 편을 추가로 찾았다. 이렇게 수집된 총 열네 편의 논문을 선행연구에서 고찰하였다. 선행연구에서 사용된 변수들은 KCYPS 2018의 범주를 따라 개인발달 영역과 발달환경 영역으로 분류하여 분석하였다.

1. 개인발달 영역

1) 학업 관련 변수

개인발달 영역에서 가장 빈번하게 사용된 변수들은 학업 관련 구인이었다(김현주

외, 2018; 문공주, 함은혜, 2016; 윤성혜, 유지원, 유영란, 2017; 임효진, 윤여경, 2017; 조선화, 2019; 천향옥, 2019). 윤성혜 외(2017)는 학업적 자기효능감이 끈기와 통계적으로 유의한 정적 관계를 맺고 있으며, 끈기를 완전매개하여 진로동기에 정적인 영향을 준다고 보고하였다. 임효진과 윤여경(2017)은 끈기를 노력지속과 흥미유지의 하위 범주로 구분하여 학업적 유능감과의 관계를 연구하였는데, 학업적 유능감 수준이 높을 때 흥미유지가 노력지속을 위하여 정적인 조절 효과를 보임을 밝혔다.

학업스트레스와 학업소진은 끈기와 통계적으로 유의한 부적 상관이 있는 것으로 보고되었다. 학업스트레스가 높으면 학업소진의 가능성이 높아지나, 끈기의 하위 범주인 노력지속 수준의 조절효과를 통하여 학업소진을 줄일 수 있었다(조선화, 2019). 같은 연구에서 끈기의 다른 하위 범주인 관심의 지속성(흥미유지)는 통계적으로 유의한 조절효과를 보이지 않았다.

문공주와 함은혜(2016)의 연구에서 과제집착력은 끈기와 정적 상관을 맺고 있었으나, 학습흥미와 끈기의 상관관계는 통계적으로 유의하지 않았다. 다만 성별에 따라 학습흥미와 끈기 간 관계가 변화하여 남자의 경우 부적, 여자의 경우 정적 상관을 갖는 것으로 나타났다.

성적의 경우, 국어, 수학, 영어 등 모든 과목의 등급을 평균한 값과 끈기 간 상관관계가 통계적으로 유의하였다(천향옥, 2019). 과제성취와 관련된 성취목표성향은 과제성향과 자아성향의 요인으로 구성된 변인으로서 과제성향만 끈기와 정적인 상관을 보였다(김현주 외, 2018).

2) 학업 이외 변수

삶의 만족도와 긍정심리자본(자기효능감, 희망, 자아탄력성, 낙관성)이 끈기와 정적으로 유의한 관계에 있었다(최효선, 양수진, 2018). 자기결정성 또한 끈기와 정적인 관계를 보였다(류재준, 임효진, 2018). 한수연과 박용한(2018a)은 자기결정동기가 끈기의 유형에 따라 차이를 보이며 끈기가 높을수록 자기결정에서의 내재적인 동기가 높다고 보고하였다. 같은 연구에서 끈기와 행동조절 변수가 정적 관계를 보였으며, 이는 자기조절능력과 끈기의 관계를 분석한 문공주와 함은혜(2016)의 연구와 일치하였다.

지능의 가변성에 대한 신념을 뜻하는 성장신념 혹은 마인드셋이 끈기와 정적 관계

가 있다는 연구가 다수 있다(류재준, 임효진, 2018; 권대훈, 2018; 한수연, 박용한, 2018b; 김현주 외, 2018). 권대훈(2018)은 성장신념과 함께 미래시간조망(future time perspective), 지연(procrastination)을 분석하였는데, 두 변수 모두 끈기와 유의한 상관을 가지는 것으로 나타났다. 류재준과 임효진(2018)은 목표유형(내재목표, 외재목표)이 성장신념과 끈기의 관계를 조절하고 있음을 보고하였다.

창의적 사고, 창의적 성향이 끈기와 정적인 상관을 보였다(임효진, 2017). 같은 연구에서 창의적 사고는 추론능력, 질문능력, 비판능력, 연합능력을, 그리고 창의적 성향은 상상력과 독창성을 측정된 것으로, 끈기가 창의적 성향을 거쳐 창의적 사고에 미치는 간접효과가 있는 것으로 나타났다.

문화예술, 취미교양, 스포츠 활동 등을 측정하는 변수인 여가활동 또한 끈기에 정적인 영향이 있는 것으로 보고되었다(최효선, 양수진, 2018). 성별의 경우, 문공주, 함은혜(2016)에서 끈기는 성별에 따른 차이가 통계적으로 유의하지 않았으나, 한수연, 박용한(2018a)의 연구에서는 끈기 수준에 따른 군집 분류를 예측하는 변수로 나타나 성별에 대한 연구 결과가 일관되지 않았다.

2. 발달환경 영역

발달환경 영역으로 가정·환경 변인, 부모와 자녀의 개방적 의사소통과 학교적응, 부모의 지지와 학업성취압력 등이 연구된 바 있다. 한수연과 박용한(2018b)은 시간 간격에 따라 가정환경과 학교환경이 끈기의 하위요소인 노력지속과 흥미유지에 미치는 영향을 검정하였다. 가정환경은 자유로운 의사소통, 다양한 경험 권장, 엄격한 규칙과 지시적 분위기로, 학교환경은 친구지지와 즐거운 반분위기, 교사의 지지와 자유로운 의사소통, 엄격한 규칙과 지시적 분위기로 측정되었다. 가정환경과 학교환경 간 설명력에는 큰 차이가 없었으며, 새롭고 다양한 경험을 권장하는 가정환경과, 친구지지 및 즐거운 반 분위기라는 학교환경이 끈기를 정적으로 관련된 것으로 나타났다.

최미경(2016)과 한정옥, 박선미(2018)의 연구 결과도 이와 유사하였는데, 끈기는 학교적응 및 부모와 자녀의 개방적 의사소통, 부모의 지지, 부모의 학업성취압력과 정적 관계를 맺고 있었다. 건강체력 또한 끈기와 정적 상관관계가 있었다. 천향욱(2019)은

고등학교에서 이루어지는 학생건강체력평가 시스템에 의거한 건강 및 체력과 관련한 변수를 분석하여 고등학생의 건강체력이 끈기와 정적 상관관계임을 밝혔다.

3. 연구방법 분석

마지막으로 본 연구에서 분석한 열네 편의 논문의 연구방법을 분석하였다. 가장 많이 사용된 연구방법은 구조방정식모형으로 여덟 편의 연구(권대훈, 2018; 류재준, 임효진, 2018; 문공주, 함은혜, 2016; 윤성혜, 유지원, 유영란, 2017; 임효진, 2017; 임효진 윤여경, 2017; 최효선, 양수진, 2018; 천항욱, 2019)에서 사용되었다. 이 때 구조방정식모형은 경로분석까지 포함하는 것으로 간주하였다. 다음으로는 중다회귀분석과 ANOVA, 위계적 회귀분석, 다항로지스틱 회귀분석 등의 회귀모형이 총 일곱 편의 연구(김현주, 김덕진, 강상범, 2018; 조선희, 2019; 한수연, 박용한, 2018b; 한수연, 박용한, 2018a; 한정욱, 박선미, 2018; 천항욱, 2019; 최미경, 2016)에서 사용되었다. 그리고 t-검정을 활용한 논문 두 편(문공주, 함은혜, 2016; 한정욱, 박선미, 2018), 군집분석을 사용한 논문(한수연, 박용한, 2018a)이 한 편 있었다.

4. 선행연구 정리

선행연구를 개인발달과 발달환경의 두 가지 영역으로 분석하였다. 대부분의 연구가 학업, 삶의 만족도, 성장신념 등과 같은 개인발달 변수를 다루며 특히 학업 관련 연구가 다수를 이루는 반면, 발달환경 변수를 분석한 연구는 많지 않았다. 또한, 모든 끈기 연구가 기존 통계기법을 활용함으로써 소수의 변수 간 관계를 통계적 검정을 통하여 설명력이 높은 모형을 찾는 것에 그쳤다. 기계학습 기법을 활용함으로써 끈기와 관련된 변수를 총체적 관점에서 조망하며 탐색하며 변수 간 관계를 파악하고 예측력이 높은 모형을 구축할 필요가 있다.

III. 연구방법

1. 자료

1) 자료 정리

KCYPS 2018은 2018년에 전국의 초4, 중1 학생들을 층화 다단계집락표집 방식으로 추출하여, 향후 7년 동안 추적 조사하는 종단연구 자료다. 본 연구에서 사용하고자는 KCYPS 2018의 중1 코호트는 해당하는 표본에 대한 추적조사의 1차년도 자료로서 생활시간, 지적발달, 진로, 사회/정서/역량 발달 등을 포함하는 개인발달 영역과 학교, 가정, 활동/문화 환경 등을 포함한 발달환경 영역의 두 가지 대영역으로 이루어져 있다.

중학교 1학년 학생 총 2,590명이 348개의 문항에 응답한 자료에 대한 데이터 클리닝 후 2,590명에 대한 302개의 변수로 정리하였다. 데이터 클리닝 과정은 다음과 같다. 먼저 일련번호, 가중치 등의 변수, 개방형 문항 변수(YMDA1B12_T, YFUR2A08_T), 분(minutes)을 묻는 변수(YTIM1A01, YTIM1B01 등)는 분석에 불필요하다고 판단하여 삭제하였다. 다음으로 거의 모든 학생들이 동일하게 답하여 분산이 0에 가까운 33개의 변수들(예: YTWIN, YTIM1G02, YMDA1A00 등)을 일괄 삭제하였다. 구조적 결측(예: 스마트폰을 사용하지 않기 때문에 사용 목적에 무응답하는 경우)에 대한 응답은 '0'으로 채코딩하였다.

2) 반응변수와 설명변수

KCYPS 2018의 끈기는 '1. 전혀 그렇지 않다' 부터 '4. 매우 그렇다'까지의 4단계 리커트 척도로 측정된 여덟 문항으로 구성되었다. 이 문항들의 평균을 반응변수로 설정하였다. 역코딩 후 문항평균은 최솟값 1.125, 최댓값, 4, 평균, 2.659, 표준편차 0.438이었고 Cronbach 알파 계수는 0.712이었다.

본 연구의 Group Mnet은 범주가 3개 이상인 변수도 분석 가능한 기법이다. 범주가 3개 이상인 변수로 장래희망 교육수준(YFUR1A00)과 시/도(학교 기준; ARA1A)가

있었다. 장래희망 교육수준(YFUR1A00)의 경우 가장 빈도가 높았던 ‘3. 4년제 대학교 졸업’을 참조집단으로 하여 4개의 더미 변수를 생성하였다. 시/도 변수의 경우 16개의 더미변수가 생성되었다(‘서울’이 참조집단). 그 외 범주가 2개인 변수 또한 더미코딩으로 처리하였고, 리커트 척도로 측정된 변수들은 각각을 연속형 변수로 취급하여 분석에 투입하였다. 그 결과 설명변수는 총 301개로 정리되었다.

3) 평가 기준

Group Mnet과 Group MCP로 중학교 1학년의 끈기 예측모형을 구축하고 평가하기 위하여 먼저 KCYPS 2018의 2,590명을 7:3의 비율로 훈련자료(training data)와 시험자료(test data)로 무작위로 분할하였다. 훈련자료로 모형적합 시 BIC(Bayesian Information Criterion)를 기준으로 조율모수(tuning parameter)를 결정하였다(식 1). 식 1의 L은 가능도, K는 추정해야하는 모수의 개수, n은 사례수를 의미한다. AIC(Akaike Information Criterion)와 더불어 모형 선택 시 주로 사용되는 지수인 BIC는 사례 수가 충분할 때 복잡한 모형에 페널티를 더 많이 부과하는 특징이 있다(Kuha, 2004).

$$BIC = -2\log(L) + K\log n \quad (1)$$

다음으로 시험자료로 각 기법의 RMSE(root mean squared error)를 구하여 Group Mnet과 Group MCP 기법을 비교하였다. 본 연구에서 훈련자료와 시험자료 분할 시 작용할 수 있는 오차를 반복적인 모형화를 통하여 통제할 목적으로 모형적합 및 예측 과정을 100번 반복하여 100개의 모형을 도출하였다. Shevade & Keerthi(2003)와 Meinshausen & Bühlmann (2010)을 참고하여 이 과정에서 선택된 변수들이 모형에서 포함된 횟수인 선택계수(selection counts) 또한 결과에 제시하였다.

2. Group MCP와 Group Mnet

기존의 회귀모형은 별점이 없는 비별점 회귀모형으로 비편향(unbiased) 추정치 중 MSE가 작은 추정치를 찾는 반면, 별점회귀모형인 Group Mnet은 추정치에 편향(bias)

을 주는 대신 분산을 줄임으로써 MSE 자체가 작아지도록 한다(유진은, 2016; Hastie, Tibshirani & Friedman, 2009). 별점 회귀모형으로 널리 쓰이고 있는 LASSO(Least Absolute Shrinkage and Selection Operator)는 회귀계수 추정치가 일치성(consistency)을 충족하지 않는 문제가 있다(Huang, Ma & Zhang, 2008; Zhao & Yu, 2006). LASSO의 볼록 별점함수(convex penalty function)가 회귀계수의 크기와 상관없이 선형으로 증가하는 반면(Huang, Ma & Zhang, 2008; Zhao & Yu, 2006; Zou, 2006), MCP(minimax concave penalty)는 작은 회귀계수에 더 큰 벌점을 주는 오목 별점함수(concave penalty function)를 이용함으로써 결과적으로 회귀계수 추정치의 일치성을 이끌어낸다(Zhang, 2010). LASSO의 다른 문제점으로 변수 간 상관이 높은 자료 분석에 적합하지 않다는 점이 있다. MCP에 ridge(능형회귀) 벌점항을 추가한 Mnet을 활용함으로써 다중공선성 문제를 다룰 수 있다(Huang, Breheny, Lee, Ma & Zhang, 2016).

정리하자면, Mnet은 LASSO의 회귀계수의 불일치성 문제를 극복할 수 있는 MCP와 다중공선성 문제를 해결할 수 있는 ridge 모형의 각 장점을 합한 방법이다. 대규모 패널자료 분석 시 다중공선성 문제에서 완전히 자유로울 수 없다고 판단하여, 본 연구에서는 Mnet을 기본 연구모형으로 선택하며 MCP와 비교하였다. 더 자세하게 설명하자면, 본 연구는 범주가 3개 이상인 변수를 모형에 포함시키기 위하여 Group Mnet과 Group MCP를 활용하였다. 일반 비별점 회귀모형에서와 범주형 변수를 더미코딩하여 집단으로 처리하듯이 별점회귀모형에서도 범주형 변수는 더미코딩 후 집단으로 모형에서의 선택 여부를 결정해야 하는데, Group Mnet과 Group MCP가 이러한 목적으로 쓰였다.

Group Mnet과 Group MCP를 식으로 설명하겠다. 설명변수가 p 개인 선형 회귀모형을 생각해 보자. 반응변수 Y 가 n 차원 벡터이고, 오차 ϵ 가 평균이 0인 n 차원 벡터일 때, 설명변수 행렬 X_k 는 k 개 범주에 대한 p_k 개 변수로 구성되는 $n \times p_k$ 행렬로 표기할 수 있다(식 2). 이 때 X_k 에 대한 회귀계수 β_k 는 $(\beta_{k,1}, \dots, \beta_{k,p_k})^T$ 로 p_k 차원 벡터가 된다.

$$Y = \sum_{k=1}^K X_k \beta_k + \epsilon. \quad (2)$$

식 3과 4는 각각 Group MCP와 Group Mnet의 목적함수(objective function)다. 식 3과 4의 우변 첫 번째 항은 OLS(ordinary least squares)의 손실함수이고, 우변 두 번째 항은 MCP의 벌점항이다. MCP 벌점항은 벌점모수 λ_1 로 회귀계수를 얼마나 축소할지 결정한다. Group MCP와 Group Mnet은 식 3과 4의 L2 norm인 $\|\beta_k\|$ 를 활용하여 범주형 변수를 집단으로 모형에서 선택하거나 선택하지 않는다. 조율모수 γ 는 회귀계수 절댓값의 크기에 따라 오목함(concavity)을 제어함으로써 벌점 속도(penalization rate)를 조절한다. 더 자세하게 설명하자면, 회귀계수의 절댓값이 두 벌점모수(γ 와 λ_1)의 곱보다 클 때 벌점이 빠르게 감소하여 회귀계수에 더 적은 벌점이 적용됨으로써 MCP와 Mnet은 LASSO보다 덜 편향된 추정치를 산출하게 된다. 식 3의 Group MCP에 ridge 벌점항을 추가하면 식 4가 된다. 식 4에서 ridge를 어느 정도로 고려할 것인지를 조율하는 벌점모수 λ_2 또한 추가되었다.

$$\hat{\beta}^{grMCP} = \operatorname{argmin}_{\beta} \left[\frac{1}{2n} \left\| Y - \sum_{k=1}^K X_k \beta_k \right\|^2 + \sum_{k=1}^K J(\|\beta_k\|, \lambda_1, Y) \right], \quad (3)$$

$$\text{where } J(x|\lambda_1, \gamma) = \begin{cases} -\frac{1}{2\gamma}x^2 + \lambda_1|x|, & |x| \leq \gamma\lambda_1. \\ \frac{1}{2}\gamma\lambda_1^2, & |x| > \gamma\lambda_1. \end{cases}$$

$$\hat{\beta}^{grMnet} = \operatorname{argmin}_{\beta} \left[\frac{1}{2n} \left\| Y - \sum_{k=1}^K X_k \beta_k \right\|^2 + \sum_{k=1}^K J(\|\beta_k\|, \lambda_1, \gamma) + \lambda_2 \sum_{k=1}^K \|\beta_k\|^2 \right]. \quad (4)$$

IV. 결 과

1. Group Mnet과 Group MCP 비교

100번의 Group Mnet과 Group MCP를 통해 도출된 RMSE를 비교하였다(표 1). Group Mnet에서 100번의 모형적합 중 평균 53.35개의 변수가 선택된 반면, Group MCP는 39.68개의 변수가 선택되었다. Group Mnet과 Group MCP는 RMSE 평균과 표준편차에 있어 거의 차이가 없었다. 수백 개의 변수로 청소년의 성장과 발달 양상을 측정하는 KCYPS 자료의 특성상 변수 간 다중공선성 문제가 발생할 수도 있다고 판단하였으므로 본 연구에서는 ridge 별점항이 추가된 Group Mnet을 분석 기법으로 선택하였다.

표 1

Group Mnet과 Group MCP 비교표

	평균 변수 개수	RMSE 평균	RMSE 표준편차
Group Mnet	53.35	0.3076	0.0068
Group MCP	39.68	0.3050	0.0068

2. Group Mnet 분석 결과

100번의 반복 후 선택계수(selection counts)를 계산하여 정리하였다(표 2). 100번 중 한 번이라도 선택된 변수 64개, 세 번 중 한 번꼴로 선택된 변수 54개, 두 번 중 한번 선택된 변수 51개, 100번 중 95번 선택된 변수 48개, 100번 모두 선택된 변수가 42개였다. 1회 이상 모형에 포함된 변수 개수와 33회 이상 모형에 포함된 변수 개수 간의 차이가 10개인데, 33회 이상, 100회 이상 모형에 포함된 변수 개수간의 차이도 12개로 반복 포함된 횟수의 간극이 커졌음에도 차이는 크게 다르지 않았다. 선택계수를 기준으로 모든 모형에서 포함된 변수들은 기존 표본이 달라진다 하여도 반응변수와 관련이 있는 변수로 포함될 것이라 판단하여 모든 모형에서 선택된 42개의 변수를

최종 변수로 선택하였다. 총 42개의 변수를 개인발달 영역과 발달환경 영역으로 나누어 표 3과 표 4에 변수의 평균, 최솟값, 최댓값을 제시하였다. 개인발달 영역에서 선택된 변수는 28개였으며, 발달환경 영역에서 선택된 변수는 14개였다.

표 2

Group Mnet 선택계수(selection counts)

반복 포함횟수	≥1	≥33	≥50	≥95	=100
변수 개수	64	54	51	48	42

1) 개인발달 영역 변수

표 3

최종 선택된 개인발달 영역 변수(28개)

변수명	설명	평균	최솟값	최댓값
YGENDER	성별	-0.0262	-0.0305	-0.0156
YTIM1J02	여가시간: 주말_운동 및 신체활동 시간	0.0068	0.0065	0.0077
YINT2A12	학업 열의: 어려운 과제도 충분히 해결할 만한 능력이 있다	0.0067	0.0040	0.0108
YINT2A16	학업 열의: 공부할 때 시간이 잘 간다	0.0375	0.0349	0.0411
YINT2B14	학업 무기력: 스스로 알아서 공부하지 않는다	-0.0380	-0.0383	-0.0331
YPSY1A04	삶의 만족도: 지금까지 내 삶에서 내가 원하는 중요한 것들을 이루어 냈다	0.0155	0.0151	0.0161
YPSY2A03	행복감: 전반적으로 아주 행복한 사람들이 있다. 나는 ...	0.0129	0.0054	0.0206
YPSY3A02	자아존중감: 때때로 나는 내가 어디에도 소용 없는 사람이라고 생각한다	-0.0329	-0.0343	-0.0301
YPSY3A03	자아존중감: 나는 내가 장점이 많다고 느낀다	0.0440	0.0408	0.0510
YPSY3A09	자아존중감: 나는 내가 실패자라고 느끼는 경향이 있다	-0.0166	-0.0229	-0.0154
YPSY4A03	주의집중: 오랫동안 집중해야 하는 과제는 하고 싶지 않다	-0.0627	-0.0678	-0.0592

변수명	설명	평균	최솟값	최댓값
YPSY4A04	주의집중: 연필이나 지우개 등 학용품을 잘 잃어버린다	-0.0223	-0.0258	-0.0177
YPSY4A05	주의집중: 주의를 기울이지 않아서 실수를 하거나 사고를 낸다	-0.0335	-0.0351	-0.0305
YPSY4A06	주의집중: 공부할 때 차분하게 앉아 있기 힘들다	-0.0473	-0.0510	-0.0456
YPSY4B02	공격성: 남이 하는 일을 방해할 때가 있다	-0.0325	-0.0338	-0.0298
YPSY4C06	신체증상: 자주 피곤하다	-0.0162	-0.0236	-0.0078
YPSY4E09	우울: 장래가 희망적이지 않은 것 같다	-0.0241	-0.0277	-0.0176
YPSY5A02	협동심: 그룹활동을 할 때, 친구가 어려워하면 기꺼이 돕는다	0.0058	0.0020	0.0093
YPSY5A05	협동심: 그룹에서 하고있는 활동이 의미있는 일이라는 것을 친구에게 이야기 한다	0.0313	0.0269	0.0346
YPSY5A14	협동심: 그룹에서 발생한 문제에 대해 잘잘못을 따지기 보다는 해결하려고 노력한다	0.0338	0.0335	0.0343
YPSY6A03	창의적 성격: 영리한	0.0149	0.0113	0.0178
YPSY6A08	창의적 성격: 가식적인	-0.0315	-0.0395	-0.0194
YPSY6A20	창의적 성격: 정직한	0.0318	0.0315	0.0331
YPSY6A25	창의적 성격: 심사숙고하는	0.0084	0.0002	0.0105
YPSY6A26	창의적 성격: 진실한	0.0394	0.0360	0.0415
YPSY6A29	창의적 성격: 자신감이 넘치는	0.0248	0.0157	0.0305
YPSY6A30	창의적 성격: 의심이 많은	-0.0407	-0.0445	-0.0330
YPHY1A00	건강상태: 건강상태 평가	0.0427	0.0374	0.0438

* 척도

YGENDER: 0(남자), 1(여자)

YTIM1J02: 1(전혀 안함), 2(30분 미만), 3(30분~1시간 미만), 4(1시간~2시간 미만), 5(2시간~3시간 미만), 6(3시간~4시간 미만), 7(4시간 이상~)

YPSY6A03 ~ YPSY6A30: 0(선택 안 함), 1(선택)

YPHY1A00: Likert- 1(전혀 건강하지 않다)~4(매우 건강하다)

위 변수 제외 모든 변수의 척도: Likert - 1(전혀 그렇지 않다)~4(매우 그렇다)

* 변수 순서는 KCYPS 설문 순서와 동일

먼저, 여학생(YGENDER)이 남학생보다 끈기가 낮았고, 여가시간 중 주말에 신체적 활동을 많이 할수록(YTIM1J02) 끈기가 높았다. 최효선과 양수진(2018)도 여가활동이 끈기에 정적인 영향을 미치는 것으로 보고하고 있다. 다만 해당 연구에서의 여가활동은 문화예술, 취미교양, 스포츠 등의 목록 중에서 가장 큰 비중의 것을 골라 빈도를 조사한 것으로, 본 연구의 주말 운동 및 신체적 활동 변수보다 범위가 넓은 것을 해석 주의할 필요가 있다.

선행연구에서 가장 빈번하게 사용된 변수들이 학업 관련 변수였던 반면, 본 연구에서는 32개의 학업관련 변수 중 단지 3개만이 선택되었다. 본 연구에서 학업 열의가 높을수록(YINT2A12, YINT2A116), 학업 무기력이 낮을수록(YINT2B14) 끈기가 높았다. 학업적 자기효능감이 끈기와 정적인 관계를 갖는다는 연구(윤성혜, 유지원, 유영란, 2017)와 학업적 유능감과 과제도전감이 끈기와 관련이 있다는 연구(임효진, 윤여경, 2017)와 일치하는 결과다.

다음으로 사회·정서·역량발달과 관련한 변수로 삶의 만족도(YPSY1A04)가 선택되었는데, 이는 최효선과 양수진(2018)이 삶의 만족도가 끈기에 정적인 영향을 미치며 직접적인 관계가 있다고 보고한 것과 그 맥락을 같이한다. 또한 본 연구에서 선택된 ‘행복감: 전반적으로 아주 행복한 사람들이 있다. 나는 ...’(YPSY2A03) 변수는 같은 선행연구의 긍정심리자본의 맥락에서 관련이 있다.

자아존중감 척도에서 세 개의 변수가 선택되었다. 선택된 변수를 살펴보면 ‘나는 내가 어디에도 소용없는 사람이라고 생각한다’(YPSY3A02)와 ‘나는 내가 실패자라고 느끼는 경향이 있다’(YPSY3A09)는 해당 변수 값이 클수록 끈기가 낮았으며, ‘나는 내가 장점이 많다고 느낀다’(YPSY3A03)는 끈기에 정적인 관계가 있는 것으로 나타났다. 선행연구에서 중·고등학생을 대상으로 자아존중감과 끈기를 연구한 논문은 없었다.

정서문제와 관련되어 주의집중은 네 개의 변수(YPSY4A03, YPSY4A04, YPSY4A05, YPSY4A06)가 선택되었는데, 모두 끈기와 부적인 관계인 것으로 나타났다. 그 이유는 해당 선택된 주의집중의 하위 문항이 모두 주의집중을 잘 하지 못하는 상황에 대하여 묻고 있었기 때문이다. 관련된 선행연구는 찾을 수 없었다. 다음으로 공격성(YPSY4B02)과 우울(YPSY4E09) 문항이 하나씩 선택되었다. 공격성과 우울은 끈기와 부적인 관계에 있는 것으로 나타났으나, 관련 선행연구는 찾기 힘들었다.

다음으로 협동심과 관련한 변수들이 세 개(YPSY5A02, YPSY5A05, YPSY5A14) 선택

되었는데, 협동심이 높을수록 끈기도 높아지는 정적인 영향을 미치는 것으로 나타났다. 마찬가지로 선행연구에서 중·고등학생을 대상으로 협동심과 끈기의 관계를 규명한 연구는 없었다.

창의적 성격과 관련한 변수가 일곱 개 선택되었다. 해당 변수들은 응답자 스스로 본인이 그렇다고 생각하면 선택하는 이분형 문항으로, ‘영리한’(YPSY6A03), ‘정직한’(YPSY6A20), ‘심사숙고하는’(YPSY6A25), ‘진실한’(YPSY6A26), ‘자신감이 넘치는’(YPSY6A29)을 선택한 학생들의 끈기가 높았고, ‘가식적인’(YPSY6A08), ‘의심이 많은’(YPSY6A30) 변수는 끈기와 부적인 관계에 있었다. 임효진(2017)의 끈기와 창의적 성향 또는 창의적 사고가 정적인 상관을 갖는다는 연구와 맥락을 같이 하는 결과다.

학생들이 자주 피곤하다(YPSY4C06)고 생각할수록, 스스로 자신의 건강상태를 건강하지 않다고 평가할수록(YPHY1A00) 끈기가 낮았다. 이는 선행연구에서의 건강체력 변수가 끈기와 정적인 상관이 있다고 보고한 천항욱(2019)의 연구와 일치한다.

2) 발달환경 영역 변수

표 4

최종 선택된 발달환경 영역 변수(14개)

변수명	설명	평균	최솟값	최댓값
YMDA1B05	사용 목적별 빈도: SNS 이용	-0.0070	-0.0087	-0.0033
YMDA1C04	의존도: 스마트폰을 사용하느라 지금 하고 있는 일(공부)에 집중이 안 된 적이 있다	-0.0284	-0.0296	-0.0238
YMDA1C05	의존도: 스마트폰 사용이 지금 하고 있는 일(공부)에 방해가 되지 않는다	0.0060	0.0050	0.0072
YMDA1C13	의존도: 스마트폰을 사용할 때 그만해야지 라고 생각은 하면서도 계속한다	-0.0468	-0.0514	-0.0457
YMDA1C15	의존도: 스마트폰 사용에 많은 시간을 보내지 않는다	0.0331	0.0304	0.0338
YACT3B02	팬덤 활동 빈도: 팬클럽이나 팬카페 활동하기	-0.0056	-0.0061	-0.0049
YFAM2D01	강요: 부모님은 나에게 늘 무엇을 하라고 하신다	-0.0017	-0.0027	-0.0003

변수명	설명	평균	최솟값	최댓값
YFAM2D03	강요: 부모님은 어떤 일을 할 때 오로지 그분들의 방식만이 유일하게 올바른 길이라고 생각하신다	-0.0081	-0.0112	-0.0037
YFAM2E01	구조제공: 내가 무언가 하려고 할 때, 부모님이 어떻게 하라고 가르쳐 주신다	-0.0113	-0.0137	-0.0056
YFAM2F02	비일관성: 부모님은 무언가를 하겠다고 말씀하시고는 실제로는 하지 않으실 때도 가끔 있다	-0.0049	-0.0062	-0.0032
YFAM2F03	비일관성: 부모님은 나에게 대한 규칙을 자주 바꾸신다	-0.0161	-0.0164	-0.0158
YEUD2A09	친구관계: 친구들과 의견 충돌이 잦다	-0.0054	-0.0057	-0.0044
YEUD2A10	친구관계: 친구와 싸우면 잘 화해하지 않는다	-0.0268	-0.0310	-0.0222
YEUD3A10	교사관계: 선생님은 내가 끈기와 인내심이 강하다고 생각하신다	0.0350	0.0330	0.0356

* 척도

YMDA1B05: Likert- 1(전혀 사용하지 않는다)~ 4(자주 사용한다)

YACT3B02: 1(전혀 없다), 2(1년에 1~2번), 3(한달에 1번), 4(한달에 2~3번), 5(1주일에 1번), 6(1주일에 여러번)

위 변수 제외 모든 변수의 척도: Likert - 1(전혀 그렇지 않다)~4(매우 그렇다)

* 변수 순서는 KCYPS 설문 순서와 동일

발달환경 영역 변수로 먼저 스마트폰과 관련한 사용 목적별 빈도(YMDA1B05), 의존도(YMDA1C04, YMDA1C05, YMDA1C13, YMDA1C15), 그리고 팬덤 활동 빈도(YACT3B02)가 선택되었다. 스마트폰 사용에 있어 SNS 이용 빈도와 의존도가 높을수록, 그리고 팬클럽이나 팬카페 활동을 자주 한다고 보고할수록 끈기가 낮았다. 스마트폰 관련 변수와 팬덤 활동 빈도 변수는 모두 선행연구에서 찾아볼 수 없었다.

양육방식의 경우 강요(YFAM2D01, YFAM2D03), 구조제공(YFAM2E01), 비일관성(YFAM2F02, YFAM2F03) 변수가 선택되었으며, 모두 끈기와 부적인 관계가 있었다. 한수연과 박용한(2018b)은 학년 초와 학년 말로 시기를 나누어 가정환경 변수와 끈기의 하위요인인 노력지속과 흥미유지 간 관계를 분석하였다. 엄격한 규칙과 지시적 분위기 요인이 노력지속과는 정적인 관계인 반면, 흥미유지의 경우 학년 초 부적인 관

계에서 학년 말 정적인 관계인 것으로 나타났다.

학교환경으로 분류되는 친구관계와 교사관계에서 세 개의 변수가 선택되었다. 친구들과 의견 충돌이 잦거나(YEUD2A09) 친구와 싸우면 잘 화해하지 않을수록(YEUD2A10) 끈기가 낮았고, ‘선생님은 내가 끈기와 인내심이 강하다고 생각하신다’(YEUD3A10) 문항과 끈기는 정적인 관계가 있었다. 이러한 친구관계와 교사관계는 최미경(2016)과 한수연과 박용환(2018b)의 연구 결과와 일치하였다.

V. 논 의

최신 기계학습 기법인 Group Mnet을 활용하여 끈기와 관련있는 변수를 총체적으로 파악하며 새로운 변수를 탐색하고 예측력이 높은 모형을 구축하는 것이 본 연구의 목적이었다. 연구 결과 선택된 42개 변수 가운데 선행연구에서 사용된 변수가 다수 있었던 반면, 아직 연구되지 않은 변수도 있었다. 즉, 수백 개의 변수를 가능한 한 모두 활용함으로써 끈기와 관련된 새로운 변수를 탐색한다는 연구 목적을 달성하였다. 주의할 점으로, 본 연구는 예측(prediction)에 초점을 맞추는 기계학습 기법을 활용하였기 때문에 결과 해석 시 설명(explanation)이 추가 되는 기존 기법과의 직접적인 비교가 적절하지 않을 수 있다. 그러나 기계학습 기법을 활용하여 청소년의 끈기와 관련된 변수를 탐색한 연구가 전무하므로, 본 절에서는 본 연구 결과를 기존 기법을 활용한 선행연구 결과와 비교하며 새롭게 탐색된 변수 및 선행연구 결과와 다른 양상을 보이는 변수에 초점을 맞추어 논하고, 후속 연구 주제를 제안하였다.

1. 본 연구에서 새롭게 탐색된 변수

선행연구에서는 사용되지 않았으나 본 연구에서 새롭게 탐색된 변수로 자아존중감, 우울, 협동심, 주의집중, 공격성, 스마트폰 관련 변수, 팬덤 활동 빈도 등이 있었다. 먼저 자아존중감과 우울의 경우 성인을 대상으로 끈기와의 관계를 보여준 연구는 있었으나(이한샘, 서은희, 2018; 이혜원, 2019), 청소년을 대상으로 하는 연구는 아직 없다. 선행연구에서 자아존중감과 우울은 끈기와 각각 정적, 부적인 관계를 보이며 본

연구 결과와 방향이 일치하였다. 협동심 또한 본 연구에서 새로이 탐색된 변수다. 협동심과 끈기를 설명변수로, 그리고 개인적 양면성을 매개변수로 하여 혁신행동에 어떠한 영향을 미치는지 분석한 연구가 있었으나(지승호, 조영복, 2017), 해당 연구에서 협동심과 끈기 간 관계는 살펴보지 않았다.

주의집중이 끈기의 하위요인인 흥미유지와 노력지속과 관련되는 구인임에도 불구하고(임효진, 2017) 끈기와의 관련성을 연구한 선행연구를 찾기 힘들었다. 공격성과 주의집중의 관계를 연구한 박영준, 조요한, 정종하, 최동호, 신성만(2016)에 따르면 낮은 주의집중력은 공격성의 수준을 직접적으로 증가시킨다. 본 연구 결과와 종합하면, 낮은 주의집중력으로 인하여 증가된 공격성이 다시 끈기를 낮출 것이라고 연구가설을 세울 수 있다. 이렇게 세 변수를 모두 투입하는 끈기 연구를 수행할 필요가 있다.

스마트폰 관련 변수와 끈기와의 관계를 분석한 연구 또한 찾을 수 없었다. 본 연구 결과, 스마트폰의 SNS 이용 빈도가 클수록, 그리고 스마트폰에 대한 의존이 높을수록 끈기가 낮아지는 부적인 관계에 있었다. 그런데 스마트폰 의존도가 높을수록 주의집중력이 떨어진다는 선행연구가 있다(박영준 외, 2016; 조은숙, 황인옥, 2018). 앞서 주의집중력이 낮을수록 공격성이 높다는 연구가 있었다. 이러한 결과를 종합하여 여러 새로운 연구 주제를 고안할 수 있다. 이를테면 스마트폰 의존도가 높을수록 주의집중력이 떨어지는데, 이를 다시 공격성 변수와 함께 투입하여 이러한 변수들이 끈기와 어떤 관계가 있는지 살펴보는 것은 흥미로운 연구 주제가 될 수 있을 것이다.

본 연구에서 팬덤활동 빈도는 끈기와 부적인 관계에 있었는데, 이 관계를 분석한 선행연구는 아직 없다. 팬덤활동이 자이존중감과 삶의 만족도에 부적인 관련이 있다는 선행연구는 있었다(최옥주, 김동성, 2018; 나재은, 2017). 스마트폰 관련 변수와 마찬가지로 향후 연구에서 팬덤 활동과 끈기와의 관계에서 자이존중감, 삶의 만족도 변수가 갖는 매개효과를 분석할 수 있을 것이다.

2. 선행연구와 다른 결과인 변수

선행연구의 결과와 본 연구의 결과가 다른 변수들을 살펴보려 한다. 그 변수들은 성별, 양육방식과 관련한 부모의 구조제공, 그리고 강요였다. 중학교 1학년을 대상으로 하는 본 연구에서는 남학생보다 여학생의 끈기가 낮을 것으로 예측하였는데, 일반

계 고등학생을 대상으로 한 문공주와 함은혜(2016)는 t-검정 후 성별에 따른 끈기 차이가 없다고 보고하였다. 이러한 차이는 연구기법뿐만 아니라 연구 대상의 차이로부터 기인했을 수 있다. 향후 끈기 연구에서 연령에 따른 성별 차이를 다룰 필요가 있다.

양육방식인 부모의 구조제공과 강요가 본 연구에서는 끈기와 부적인 관계였는데, 한수연과 박용한(2018b)의 연구에서는 그 영향이 시기에 따라 다르게 나타났다. 해당 연구에서도 밝히고 있듯이 끈기에 있어 학년 초와 학년 말의 차이가 크지 않았다. 따라서 학년 초와 학년 말의 개인 내에서의 끈기가 유의한 차이가 있는지를 밝힌 후 양육방식에서의 구조제공과 강요가 끈기와 갖는 관계를 후속 연구에서 더 명확히 규명할 필요가 있다.

3. 연구 의의 및 제언

끈기가 노력과 결부되며 성취에 절대적인 영향을 끼칠 것이라는 주장이 득세하고 있다. 특히 끈기가 학업뿐만 아니라 진로 및 결혼과 같은 삶의 전반에 걸친 다양한 장면에서 성취와 성공을 예측하는 변수로 밝혀지고 있는 만큼, 청소년의 끈기 수준을 향상시키기 위한 교육적 개입이 요구된다. 그러나 그간 청소년을 대상으로 끈기를 측정하는 대용량/패널자료가 없어서 관련 연구가 그다지 많지 않았으며, 특히 기계학습법을 활용하여 분석한 연구는 찾을 수 없었다. 2018년부터 한국청소년정책연구원이 통계청 승인통계인 KCYPS 조사 항목으로 끈기를 포함시킴으로써 앞으로 관련 연구가 활발할 것으로 기대된다. 본 연구는 KCYPS 2018 원년 자료에 기계학습 기법을 활용하여 청소년의 끈기와 관련된 변수를 포괄적으로 탐색하였다. 본 연구 결과가 끈기와 다른 변수간 관계를 규명하는 향후 연구에서 변수 선택 시 이론적 근거를 제공할 수 있을 것으로 기대한다. 본 연구의 의의를 다음과 같이 몇 가지 실제적·학술적 의의로 정리하였다.

첫째, 본 연구에서는 KCYPS 2018이 제공하는 수백 개에 이르는 변수들을 기계학습 기법을 이용하여 하나의 모형에 투입하여 살펴보았다. 기존의 끈기와 관련된 연구에서는 이론에 근거하여 한정된 변수나 요인들로 끈기와의 관계를 모형화한 반면, 본 연구에서는 이론에 구애받지 않고 관련이 없어 보일 수 있는 변수들까지 모두 모형에

투입하여 끈기와 관련된 새로운 변수를 파악하고자 하였다. 기존의 연구방법으로는 이처럼 수많은 변수를 모형에 동시에 투입하여 종합적으로 고려하는 것이 쉽지 않다. 기계학습 기법 중 하나인 벌점회귀모형을 활용함으로써 연구기법 상 제한점을 해결하고 끈기와 관련되는 여러 새로운 변수를 파악하며 다양한 향후 연구 주제를 제시할 수 있었다는 점이 본 연구의 첫 번째 의의라고 할 수 있다.

둘째, 연구 결과 흥미로운 점으로 본 연구에서 학업 관련 변수가 거의 선택되지 않았다는 것을 들 수 있다. 기존 연구에서 학업 관련 변수가 가장 빈번하게 사용된 반면, 본 연구 결과 학업 성취 만족도와 학업 열의, 학업 무기력을 측정하는 총 32개의 학업관련 변수 중 단 3개 변수만이 선택되었다. 최종 모형에서 개인의 심리적 특성, 양육방식과 학교환경 등의 변수가 오히려 더 많이 선택되었다는 점으로 미루어볼 때, 끈기와 학업 관련 변수에 관한 심도 있는 이론적 고찰이 필요하다. 향후 연구에서는 끈기 관련 연구를 학습 상황에 국한시키지 않고 더 넓은 의미의 학생 생활 전반에 걸친 변수들을 활용하며 탐색할 필요가 있는 것으로 보인다.

셋째, 본 연구는 Group Mnet을 활용함으로써 Group LASSO으로 KCYPS 자료를 분석한 유진은과 노민정(2017)의 연구에서 진일보하였다. 즉, Group Mnet을 적용하여 Group LASSO의 단점인 회귀계수의 불일치성을 극복하고 다중공선성 문제를 다룰 수 있었으며, 범주가 3개 이상인 변수 또한 처리할 수 있었다는 점에서 학술적인 의의가 있다. 또한, 벌점회귀모형 구축 시 자료 분할의 임의성을 극복하기 위하여 선택 계수(selection counts) 기법을 연구에 적용한 것에서도 학술적 의의를 찾을 수 있다.

넷째, 본 연구에서 활용한 기계학습 기법인 벌점회귀모형의 장점을 언급하고자 한다. 예측을 중시하는 기계학습 기법은 설명을 중시하는 기존 기법과 패러다임부터 다르다. 기계학습 기법들은 설명력보다 예측력을 우선한다. 소위 블랙박스 모형을 도출한다고 알려진 신경망(neural network), 랜덤 포레스트(random forest), 서포트벡터기계(support vector machines) 등이 특히 그러하다. 이러한 기법들은 새로운 자료에 대한 예측은 잘 하지만, 왜, 어떻게 그러한 결과가 도출되는지를 잘 설명하지 못하며, 모형 도출 과정을 이해하기도 어렵다. 블랙박스 모형 기법과 비교할 때, 선형 회귀모형에 기반한 벌점회귀모형은 예측력을 고려하면서도 상대적으로 해석하기 쉽고 설명하기 좋은 모형을 도출한다는 장점이 있다.

다섯째, 넷째와 관련하여 벌점회귀모형은 통계적으로 유의한 회귀계수를 선택하는

것이 아닌, 모형 예측력을 최대화 하는 방향으로 모형의 선택이 이루어진다(T. Hastie, personal communication, Jan. 28, 2017). 본 연구 결과 선택된 42개의 변수는 단순히 모형 구축에 활용된 KCYPS 2018 자료를 잘 설명하는 것에 그치는 것이 아니라, 모형 구축 시 활용되지 않은 새로운 자료를 잘 예측할 수 있는 변수들이다. 이와 관련하여, 중학생의 끈기와 관련이 있는 변수를 탐색한 본 연구는 이론에 근거한 통계적 유의성 검증보다는 반응변수에 대한 총체적 관점에서의 탐색 및 예측모형 구축에 더 초점이 맞춰져 있다. 따라서 본 연구에서는 상관계수 또는 결정계수를 보고하는 대신, 유진은과 노민정(2017)의 연구에서와 같이 기법 간 예측오차를 비교하였다. 본 연구를 기반으로 앞으로 기계학습 기법을 활용하는 연구가 활발해져서 조만간 끈기에 대한 예측모형 간 비교가 가능하게 될 것을 기대한다.

참고문헌

- 권대훈 (2018). 성장 마인드셋, 미래시간조망, 그릿과 지연의 관계. **아시아교육연구**, 19(3), 725-744. doi:10.15753/aje.2018.09.19.3.725
- 김현주, 김덕진, 강상범 (2018). 청소년 검도선수의 마인드셋과 그릿의 관계에서 성취 목표성향의 매개효과. **대한무도학회지**, 20(4), 47-60.
- 김희명, 황매향 (2015). 한국판 아동용 끈기(Grit) 척도의 타당화. **교육논총**, 35(3), 63-74. doi:10.25020/je.2015.35.3.63
- 나재은 (2017). 초기청소년의 팬덤활동이 삶만족도에 미치는 영향: 자아존중감의 조절 효과. **한국콘텐츠학회논문지**, 17(8), 396-405. doi:10.5392/JKCA.2017.17.08.396
- 류재준, 임효진 (2018). 성장신념, 자기결정성, 그릿의 관계에서 목표 유형의 조절효과. **교육심리연구**, 32(3), 397-419. doi:10.17286/KJEP.2018.32.3.04
- 문공주, 함은혜 (2016). 고등학생의 투지, 흥미, 과제집착력, 자기조절능력 및 과학학업성취의 관계 분석. **한국과학교육학회지**, 36(3), 445-455. doi:10.14697/jkase.2016.36.3.0445
- 박영준, 조요한, 정종하, 최동호, 신성만 (2016). 스마트폰 의존과 공격성 간의 관계에서 주의집중의 매개효과. **한국심리학회 학술대회 자료집**, 407-407.
- 유진은 (2016). 데이터 마이닝 기법을 통한 교육 패널데이터 분석: 별점회귀모형과 KYPS 자료. **아시아교육연구**, 17(3), 1-19. doi:10.15753/aje.2016.09.17.3.1
- 유진은, 노민정 (2017). Group lasso를 통한 중학생의 삶의 만족도에 영향을 미치는 변수 탐색. **한국청소년연구**, 28(1), 127-149. doi:10.14816/sky.2017.28.1.127
- 유진은 (2019). 기계학습: 대용량/패널자료와 학습분석학 자료 분석으로의 활용. **교육공학연구**, 35(2), 313-338. doi:10.17232/KSET.35.2.313
- 윤성혜, 유지원, 유영란 (2017). 고등학생의 학업적 자기효능감과 진로동기 간의 관계에서 그릿(grit)의 매개효과 검증. **직업교육연구**, 36(3), 47-66.
- 이한샘, 서은희 (2018). 대학생의 우울과 자아탄력성의 관계에서 자기효능감과 그릿의 매개효과. **교육학연구**, 56(3), 61-88. doi:10.30916/KERA.56.3.61
- 이혜원 (2019). 예비유아교사의 자아존중감과 교직선택동기 간의 관계에서 그릿의 매개효과. **교육의이론과실천**, 24(1), 21-36. doi.org/10.26804/kdge.2019.24.1.21

- 임효진 (2017). 그릿(Grit)과 창의적 성향, 창의적 사고의 구조적 관계. **사고개발**, 13(2), 45-65.
- 임효진, 윤여정 (2017). 그릿(Grit)과 과제도전감의 관계: 흥미유지와 노력지속에서 유능감의 조절효과. **교육방법연구**, 29(4), 657-675. doi:10.17927/tkjems.2017.29.4.657
- 조선화 (2019). 중학생의 학업스트레스와 학업소진과의 관계에서 그릿의 조절효과. **청소년문화포럼**, 57, 145-169. doi:10.17854/ffyc.2019.01.57.145
- 조은숙, 황인옥 (2018). 초등학생의 스마트폰 의존과 주의집중력의 관계에서 자아탄력성의 매개효과. **한국콘텐츠학회논문지**, 18(6), 131-143. doi:10.5392/JKCA.2018.18.06.131
- 지승호, 조영복 (2017). 개인적 양면성의 선행요인과 결과요인에 관한 연구: 개인적 양면성의 매개효과를 중심으로. **한국인사관리학회 학술대회 발표논문집**, 3, 1-25.
- 최미경 (2016). 부모와의 의사소통이 남고생의 학교적응에 미치는 영향: 투지(Grit)의 매개효과. **청소년학연구**, 23(3), 75-102. doi:10.21509/KJYS.2016.03.23.3.75
- 최옥주, 김동성 (2018). 청소년의 팬덤활동이 자아존중감에 미치는 영향: 자아탄력성의 매개효과. **디지털융복합연구**, 16(10), 243-251. doi:10.14400/JDC.2018.16.10.243
- 천항욱 (2019). 남자고등학생의 건강체력, 학업성취, Grit의 관계. **한국체육학회지**, 58(1), 133-146. doi:10.23949/kjpe.2019.01.58.1.10
- 최효선, 양수진 (2018). 고등학생의 여가활동과 삶의 만족도의 관계에서 그릿과 긍정심리자본의 매개효과. **한국심리학회지: 발달**, 31(2), 105-123.
- 한수연, 박용한 (2018a). 중학생의 그릿 유형에 따른 행동조절과 자기결정동기의 차이. **교육방법연구**, 30(1), 73-101. doi:10.17927/tkjems.2018.30.1.73
- 한수연, 박용한 (2018b). 중학생의 그릿 변화와 관련된 가정·학교 환경요인 탐색. **열린교육연구**, 26(4), 175-201. doi:10.18230/tjye.2018.26.4.175
- 한정욱, 박선미 (2018). 중학생이 인지한 부모 지지와 학업성취압력이 그릿(Grit)에 미치는 영향. **학습자중심교과교육연구**, 18(20), 165-182. doi:10.22251/jlcci.2018.18.20.165
- Duckworth, A. (2016). *Grit: The power of passion and perseverance*. New York: Scribner/Simon & Schuster.

- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long term goals. *Journal of Personality and Social Psychology*, 92(6), 1087-1101. doi:10.1037/0022-3514.92.6.1087
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Elements of statistical learning* (2nd ed.). New York: Springer.
- Huang, J., Ma, S., & Zhang, C. H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica*, 3, 1603-1618.
- Huang, J., Breheny, P., Lee, S., Ma, S., & Zhang, C. H. (2016). The Mnet method for variable selection. *Statistica Sinica*, 26, 903-923. doi:10.5705/ss.202014.0011
- Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research*, 33(2), 188-229. doi:10.1177/0049124103262065
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Methodology)*, 72(4), 417-473. doi:10.1111/j.1467-9868.2010.00740.x
- Shevade, S. K., & Keerthi, S. S. (2003). A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics*, 19, 2246-2253. doi:10.1093/bioinformatics/btg308
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100-1122. doi:10.1177/1745691617693393
- Yoo, J. E. (2018). TIMSS 2011 student and teacher predictors for mathematics achievement explored and identified via elastic net. *Frontiers in Psychology*, 9, 317. doi:10.3389/fpsyg.2018.00317
- Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2), 894-942. doi:10.1214/09-AOS729
- Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7, 2541-2563.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429. doi:10.1198/016214506000000735

ABSTRACT

An exploration of the variables relating to middle school students' grit via a machine learning technique, Group Mnet

Yoo, Jin Eun* · Kim, Hyeong Gwan* · Rho, Minjeong*

Previous studies on adolescents' grit focused on testing the relationships of a small number of variables based on literature reviews, utilizing traditional statistical techniques. The purpose of this study was to explore the hundreds of KCYPS (Korean Children and Youth Panel Survey) 2018 variables and to construct a prediction model via a machine learning technique, group Mnet. Mnet is categorized as penalized regression among machine learning techniques, and deals with the inconsistency and multicollinearity problems of LASSO and MCP, respectively. Group Mnet also handles categorical variables. Specifically, this study repeated 100 times of model building after random data-splitting, and 42 variables were selected out of 301 after selection counts. Variables measuring academic lethargy, life satisfaction, happiness, creative personality, health status, parenting style, gender, and relationships with friends as well as teachers had been investigated in previous research, and were also selected in this study. Newly found variables included predictors relating to self-esteem, cooperation, depression, attention, aggression, smartphone usage, and fandom activities. The significance and implications of the study as well as future research topics were discussed.

Key Words: grit, penalized regression model, machine learning, Group Mnet, Korean Youth Panel Survey(KCYPS)

투고일: 2019. 12. 9, 심사일: 2020. 1. 31, 심사완료일: 2020. 2. 13

* Korea National University of Education