

Artificial Intelligence-Based High School Course and University Major Recommendation System for Course-Related Career Exploration

Jinheon Baek[†] · Hayeon Kim^{**} · Kiwon Kwon^{***}

ABSTRACT

Recent advances in the 4th Industrial Revolution have accelerated the change of the working environment, such that the paradigm of education has been shifted in accordance with career education including the free semester system and the high school credit system. While the purpose of those systems is students' self-motivated career exploration, educational limitations for teachers and students exist due to the rapid change of the information on education. Also, education technology research to tackle these limitations is relatively insufficient. To this end, this study first defines three requirements that education technologies for the career education system should consider. Then, through data-driven artificial intelligence technology, this study proposes a data system and an artificial intelligence recommendation model that incorporates the topics for career exploration, courses, and majors in one scheme. Finally, this study demonstrates that the set-based artificial intelligence model shows satisfactory performances on recommending career education contents such as courses and majors, and further confirms that the actual application of this system in the educational field is acceptable.

Keywords : Artificial Intelligence, EduTech, Career Exploration Recommendation System

교과 연계 진로 탐색을 위한 인공지능 기반 고교 선택교과 및 대학 학과 추천 시스템

백진현[†] · 김하연^{**} · 권기원^{***}

요약

4차 산업 혁명 시대의 도래에 따라 직업 환경의 변화가 가속화되고 있으며, 이와 함께 교육의 패러다임이 자유학기제와 고교학점제에 바탕을 둔 진로교육을 중심으로 변화하고 있다. 하지만, 학생들의 자율적인 진로 탐색을 지향하는 자유학기제 및 고교학점제의 정책적 목표와 달리, 진로교육 콘텐츠의 개발과 이용에 있어 교사 및 학생들의 한계가 존재하고, 이를 뒷받침할 에듀테크 기술 연구 역시 상대적으로 부족한 실정이다. 따라서 본 연구는, 교육 현장에서의 진로교육 실태를 바탕으로, 에듀테크 기술이 교과연계 진로교육과 관련해 갖춰야 할 요구조건을 세 가지로 정의하였다. 다음으로 데이터 기반 인공지능 기술을 통해, 진로탐색용 탐구주제와 고교 과목, 그리고 대학에서 수학 가능한 전공을 아우를 수 있는 데이터 시스템 및 인공지능 추천 모델을 제안하였다. 마지막으로 실험을 통해, 셋 인코딩-디코딩 기반 인공지능 추천 모델이 진로교육 콘텐츠 추천에서 만족할 만한 성능을 보이는 것을 확인하였고, 교육 현장에서의 실제 적용 결과 또한 만족스럽다는 것을 확인하였다.

키워드 : 인공지능, 에듀테크, 진로 탐색, 추천 시스템

1. 서론

4차 산업 혁명의 시대가 도래하면서 기존의 직업이 사라지고 새로운 직업이 생겨나는 속도가 과거에 비해 빨라지고 있다[1]. 이와 함께 국내 교육의 패러다임 또한 진로교육을 중심으로 변화하는 중이다. 그에 대한 대표적인 예로, 중학교

수준에서는 학생의 다양한 체험 활동 및 진로 탐색을 장려하는 자유학기제가 2013년 전국 42개교를 시작으로 2016년 전면 실시되었으며, 2020년 이것이 확대된 자유학년제가 전면 시행되고 있다[2, 3]. 또한, 고등학교 수준에서는 학생의 과목 선택권을 확대함으로써 진로 설계를 지원하는 고교학점제가 2018년 105개 학교를 시작으로 부분 시행 중이며, 2025년 전국 고등학교에 전면 도입될 예정이다[4].

위와 같은 자유학년제와 고교학점제는 문·이과 통합형 교육과정을 표방하는 2015 개정 교육과정의 기반 위에, 진로 탐색 및 설계에서의 학생 선택권을 확대하고자 하는 목표를 가지고 있다. 이는 평생직장의 개념이 사라지고 있는 4차 산업 혁명 시대에, 문·이과를 초월한 폭넓은 학습, 다양하고 구

* 이 논문은 2020년도 서울시 산학연 협력사업(IC190048)의 재원으로 서울R&D지원센터의 지원을 받아 수행된 연구임.

[†] 준 회 원 : 한국과학기술원 인공지능대학원 석사과정

^{**} 준 회 원 : 서울대학교 국어교육과 박사과정

^{***} 준 회 원 : 한국방송통신대학교 이터닝학과 석사과정

Manuscript Received : October 15, 2020

Accepted : November 5, 2020

* Corresponding Author : Hayeon Kim(khy3221@snu.ac.kr)

체적인 진로 체험을 바탕으로 학생들의 진로 개척 역량을 강화하기 위함이다. 이를 통해 변하는 직업 환경에 유연하게 적응할 수 있는 창의적·융합적 인재를 양성하고자 한다.

자유학기제와 고교학점제의 지향점은 학생의 선택권을 보장함으로써 개별 학생의 관심사와 적성, 다양성을 존중하는 맞춤형 진로교육이다. 따라서 실질적인 맞춤형 진로교육이 실현되기 위해서는 먼저 학생들의 폭넓은 학습과 체험을 지원할 수 있는 다양한 진로·진학 교육 콘텐츠가 마련되어야 하며, 풍부한 콘텐츠 자원을 바탕으로 교사가 개별 학생의 진로 탐색 및 설계를 효과적·효율적으로 도울 수 있는 학습 환경이 마련되어야 한다. 또한, 교과연계 진로교육이 교과목표와 진로목표를 두루 달성하는 한편 학생과 교사의 효능감까지도 제고 한다는 연구 결과[5]를 참고할 때, 진로교육을 포괄하는 교과교육 콘텐츠를 마련하여 진로교육과 교과교육의 학습효과를 동시에 꾀할 필요가 있다.

그러나 학생들의 자율적인 진로탐색을 지원한다는 근원적 목표와는 달리, 1) 교사들이 모든 교과에 적합한 모든 진로 분야를 알 수 없고, 2) 진로탐색 활동에 활용될 다양하고 창의적인 진로교육 콘텐츠가 부족한 상황이며[6], 3) 교사들 또한 교육 콘텐츠 생성에 어려움을 겪고 있어[7] 해당 정책의 실질적인 목표 달성에 의문이 제기되고 있다. 이와 관련해 데이터 기반 인공지능 등의 정보통신기술을 활용한 차세대 에듀테크를 바탕으로 1) 학생의 진로 탐색 및 설계를 효과적·효율적으로 돕고, 2) 진로교육과 교과교육을 연계하는 콘텐츠를 마련할 수 있으나, 이를 뒷받침 할 데이터와 기술적 연구 역시 상대적으로 부족한 실정이다.

그렇다면 양질의 다양한 교과연계 진로교육 콘텐츠를 차세대 에듀테크 기술을 바탕으로 갖추는 것이, 교사의 원활한 진로지도도를 도우면서도 다양한 학생의 요구를 충족시킬 수 있는 제도권 내 맞춤형 진로교육의 핵심이라 할 수 있다. 이에 따라 본 연구는 앞서 설명한 맞춤형 진로교육의 핵심 목표를 달성하기 위해 첫째, 전국의 고등학교 교사 8명과 고등학생 1,527명에 대한 인터뷰와 설문조사를 바탕으로, 실제 진로교육 현장에서 필요한 데이터 시스템 개발 요건을 정의한다. 둘째, 정의된 시스템 개발 요건을 바탕으로, 교과연계 진로교육 콘텐츠를 효과적·효율적으로 제공하기 위한 인공지능 모델을 개발한다. 인공지능 모델은 교과 및 진로 콘텐츠를 입력으로 받아, 이를 셋 기반으로 벡터 공간에 임베딩 한 다음[8], 학생들이 선호할 만한 교과연계 진로 정보를 추천하는 모델이다. 마지막으로 인공지능 모델이 제공하는 콘텐츠 추천 정확도에 대한 정량적·정성적 성능과 함께 실제 교육 현장에서의 만족도를 확인한다.

본 연구의 기여를 세 가지로 요약하면 다음과 같다.

- 교사 대상 현장 인터뷰와 학생 대상 설문조사를 바탕으로, 실제 진로교육 현장에서 요구되는 교과연계 진로교육 콘텐츠의 요건을 정의한다.
- 현장 조사를 통해 정의한 시스템 개발 요건을 바탕으로, 맞춤형으로 교과연계 진로 콘텐츠를 개발 및 제공할 수 있는 데이터 시스템 및 인공지능 모델을 제안한다.
- 콘텐츠 추천에 대하여, 제안한 데이터 시스템 및 인공지능 모델의 정량적·정성적 성능을 확인하고, 사용 측면에서 실제 교육 현장에서의 결과를 확인한다.

2. 관련 연구

2.1 진로교육 패러다임의 변화

자유학기제란 중학교에서 운영되고 있는 제도로서, 한 학기 동안 학생들이 시험 부담에서 벗어나 꿈과 적성을 찾을 수 있도록 교육과정을 유연하게 운영하는 제도를 일컫는다[9]. 대표적으로 해당 제도 내에서, 토론·실습 등 학생 참여형으로 수업을 개선하고, 진로탐색 활동과 같은 체험 활동을 실시하고 있다. 더불어 고등학교에서 운영되는 고교학점제란, 진로와 적성에 따라 다양한 과목을 선택 이수하여 누적 학점이 기준에 도달하면 졸업을 인정받는 제도이다[4].

현재 자유학기제가 확대된 자유학년제가 중학교에서 전면 시행 중이며, 고교학점제는 2025년 전국 고등학교에 전면 도입될 예정이다. 학기 단위에서 학년 단위로 확대된 자유학년제, 일부 학교에서 전국 학교로 확대되어가는 고교학점제는 학생의 선택권을 강화함으로써 폭넓은 진로 체험과 학습을 도모하려는 교육정책의 방향성을 보여주고 있다.

만족도 측면에서, 자유학기제가 성적향상에 긍정적 영향을 미칠 수 있고[10], 고교학점제가 학생의 과목 선택권 확대에 기여하는 등[11] 여러 긍정적 측면이 확인되고 있으나, 아직은 과도기적 단계로서 문제점 또한 제기되고 있다. 대표적으로 제도의 안착과 목적 달성을 위해서는 교사들의 제도에 대한 이해, 학습 컨설턴트로서의 역량 제고가 필요하다는 의견이 제기되고 있다[12, 13].

2.2 교과연계에 기반을 둔 진로교육

학생들의 만족도와 교사의 자신감을 동시에 높일 수 있다는 연구[5]와 고등교육에서 전공 맞춤형으로 학습 콘텐츠를 운영할 때 학습 효과가 크게 상승하였다는 연구[14]를 종합할 때, 교과연계 진로교육은 교과교육과 진로교육의 동시 상승효과를 시사한다. 따라서, 교과연계 진로 콘텐츠의 개발이 중요하며, 또 이를 현장의 교사 및 학생들이 쉽게 이용할 수 있도록 하는 접근성 역시 중요하다.

2.3 인공지능 기술을 활용한 교육 서비스

교육 분야 인공지능 관련 연구는 1985년 1편을 시작으로, 2016년 60편 이상의 논문이 게재되는 등 급격하게 증가하고 있다[15]. 그러나 국내의 관련 연구는 AI를 교과 교육에 직접 적용하는 연구가 아닌, 교육 환경 개선에 적용하는 연구가 대부분이다[16]. 이는 교과교육에서의 AI 적용 방안을 탐구하는 본 연구의 필요성을 뒷받침한다.

2.4 교육 콘텐츠 개발을 위한 셋 표현 인공지능

본 연구는 교과연계 진로교육 콘텐츠의 개발을 위해, 교과 콘텐츠(선택교과 내 단위)와 진로 콘텐츠(세부전공), 그리고 이를 아우르는 교과 연계 진로 콘텐츠의 유기적인 연결점을 인공지능 기술을 통해 찾고자 한다. 이러한 인공지능 기술을 구현하기 위해서는, 교과 콘텐츠와 진로 콘텐츠를 임베딩 공간에서 벡터 표현으로 잘 표현할 수 있어야 한다.

먼저 두 콘텐츠의 특징을 살펴보면 다음과 같다. 1) Varying Cardinality: 교과 연계 진로 탐구 주제와 관련된 교과 콘텐츠와 진로 콘텐츠의 개수는 N개로 고정하여 정의할 수 없으며, 그 길이가 탐구 주제에 따라 가변한다. 2) Permutation Invariance: 교과 연계 진로 탐구 주제와 관련된 교과 콘텐츠와 진로 콘텐츠의 나열 순서에 따라, 탐구 주제가 지향하는 궁극적 목표는 변하지 않는다. 따라서 두 특성을 모두 만족하는, 콘텐츠 개발 및 추천을 위한 인공지능 모델로서 본 연구에서는 셋 표현 모델을 사용한다.

기존의 셋 표현 인공지능 모델은 주로 3D shape recognition [17, 18], generating point clouds[19, 20], few-shot image classification[21, 22] 문제 등에서 사용되었다. 또한 셋 표현 인공지능을 위한 이론적인 배경을 제공한 연구도 존재한다[17]. 더 나아가, 자연어 이해 문제에서 높은 성능을 보인 Transformer[23] 아키텍처 기반으로 셋 표현을 생성하는 연구도 존재한다[24].

3. 교육 현장 조사를 기반으로 한 진로탐색 시스템 개발 요건의 정의

본 장에서는 에듀테크 기반 교과연계 진로교육 콘텐츠를 개발 및 추천하는 인공지능 모델을 제안하기에 앞서, 진로교육 실태조사를 바탕으로 실제 학교 현장에서 교사와 학생들이 겪고 있는 어려움을 확인한다. 그다음, 도출된 학교 현장에서의 문제점을 바탕으로 데이터 시스템과 인공지능 모델이 가져야 할 시스템 개발 요건을 세 가지로 정의한다.

3.1 교육 현장 조사

시스템 제안에 앞서, 실제 진로교육 현장에서 교사와 학생들이 겪고 있는 교과연계 진로교육의 어려움을 파악하고자 인터뷰와 설문조사를 실시하였다.

교사 대상 현장 조사 먼저 학교 현장에서 진로·진학 지도 경험이 있는 국내 고등학교 교사 8명을 대상으로 교육 방법 및 콘텐츠 개발에 대한 심층 인터뷰를 하였다. 교사 한 명당 약 20분의 시간이 소요되었으며, 인터뷰 결과 교사들이 진로·진학 지도에서 겪는 공통적 어려움 혹은 문제점은 크게 세 가지로 정리되었다.

첫째, 국내에 존재하는 모든 학과 및 직업 정보를 한 명의 교사가 섭렵할 수 없기에, 학생들의 다양한 요구를 맞춰주기가 어렵다. 이로 인해 특정 교과목과 관련한 진로 정보가 필요한 학생이 있더라도 커리어넷, 대학 알리미 등의 진로·진학 정보 링크를 알려주거나, '희망하는 학과 사이트에 들어가서 정보를 찾아봐라' 정도의 답변만을 해줄 때가 많다. 이처럼 진로 캠프나 강연 등 공통 진로활동을 제외하면 사실상 교과연계 진로교육은 개별 학생의 자율성에 의존하고 있다.

둘째, 교과수업에서 학생이 진로탐색을 위한 탐구활동을 실시하더라도, 전문성의 부족으로 명확한 탐구 가이드라인을 제공하기 어렵다. 교사의 전공과 관련된 내용은, 오류를 짚어

주거나 탐구 내용의 발전을 위한 피드백을 주기가 비교적 쉽지만, 교사의 전공과 먼 내용이거나 전문적인 내용은 학생의 탐구활동을 돕기가 어렵다. 특히 현대에는 학문이나 기술의 발전 속도가 매우 빠르고 정보도 넘치기 때문에, 전문적이면서도 정확한 피드백을 주기 어려운 경우가 많다.

셋째, 개별 학생의 흥미와 관심사를 고려한 진로교육 콘텐츠를 개발하기 어렵다. 교과연계 진로교육 콘텐츠의 경우 더욱 그렇다. 각종 행정업무와 교과수업만으로도 업무가 과중하기에, 다양하고, 새로우며, 정확한 진로교육 콘텐츠를 개발할 만한 시간적 여유가 부족하다. 특히 연차가 오래된 교사의 경우 디지털 기기 사용에 익숙하지 않기 때문에, 원격수업 등 변화된 학습 환경에 적합한 콘텐츠를 개발하는 데 더 어려움을 겪고 있다.

학생 대상 현장 조사 다음으로, 전국 고등학생 1,645명을 무작위로 추출, 교과수업에서의 진로 및 진학 탐색 활동에서 겪고 있는 어려움에 대하여 설문조사를 실시하였다. 응답자는 1학년 588명, 2학년 691명, 3학년 366명으로 구성되었으며, 이 중 1,419명(86.3%)이 일반 인문계고등학교 학생이다. 설문지는 총 5문항이며, 이분형 질문, 체크리스트형 질문, 서열형 질문 등을 활용하였다. 각 질문에 대한 응답 결과를 분석하면 아래와 같다. 자유 서술 항목인 5번 문항의 경우, 응답의 경향성을 전체적으로 살펴보고 키워드 추출 후, 워드클라우드를 시각화하였다(Fig. 5).

1번 문항(Fig. 1)을 살펴보면, 교과수업에서 진로탐색 활동 시 정보 부족으로 어려움을 겪었던 학생이 무려 75.2%에 달한다. 희망하는 진로에 대한 정보뿐만 아니라, 진로와 교과수업의 관련성을 찾는 데에도 어려움을 느낀다는 사실을 확인할 수 있다. 이는 대학 진입을 위한 전공 선택과 밀접한 관련이 있기에, 장기적인 관점에서 진로 설계를 돕기 위해서는 양질의 교과연계 진로교육 콘텐츠가 개발될 필요가 있다.

1. 교과수업에서 진로탐색 활동 시 정보 부족으로 어려움을 겪은 적이 있나요(Fig. 1)?

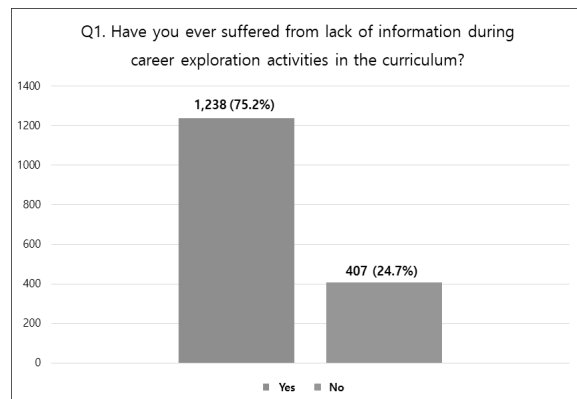


Fig. 1. Response Result of Question 1

2. 교과수업에서 진로탐색 활동 시 누구의 도움을 받고 있나요(Fig. 2)? (복수 응답 가능)

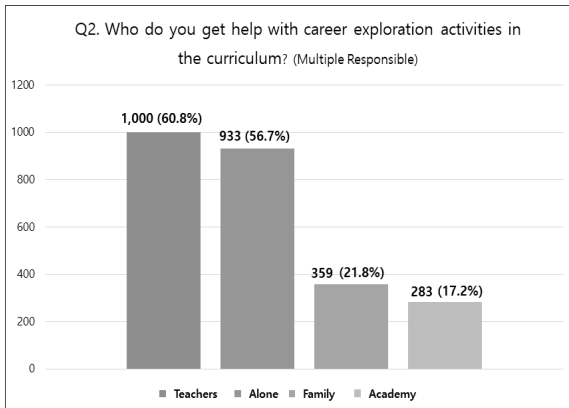


Fig. 2. Response Result of Question 2

3. 교과수업에서 진로탐색 활동 시 가장 많이 참고하는 자료는 무엇인가요(Fig. 3)? (복수 응답 가능)

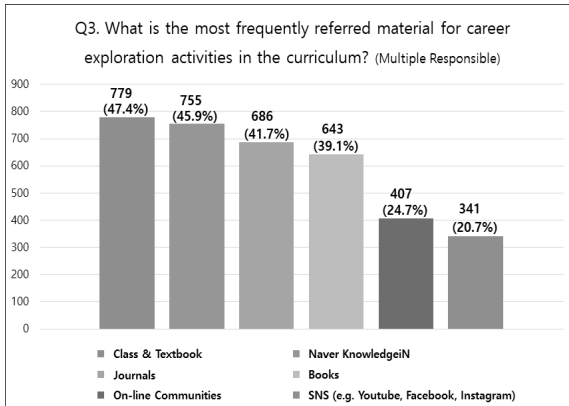


Fig. 3. Response Result of Question 3

4. 교과수업에서 진로탐색 활동을 준비할 때 어떤 것이 가장 어렵나요(Fig. 4)?

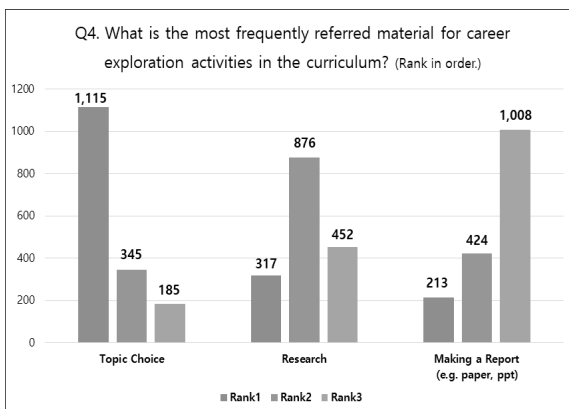


Fig. 4. Response Result of Question 4

5. 교과수업에서 진로탐색 활동을 준비할 때 기타 어려운 점이 있다면 자유롭게 작성해 주세요(Fig. 5).



Fig. 5. Wordcloud: Response Result of Question 5

2, 3번 문항(Fig. 2, Fig. 3)에서는 교과연계 진로교육이 대부분 학교 선생님 또는 수업, 그리고 학생의 자발적 검색 활동에 의존하고 있다는 사실이 나타난다. 그러나 앞서 교사 인터뷰에서 확인한 것처럼, 교사가 다양한 진로 정보를 파악하는 데에는 한계가 있다.

주제 선정 및 자료조사에 어려움을 느끼고 '무엇을 어떻게 해야 할지' 모르겠다는 4, 5번 문항(Fig. 4, Fig. 5)의 결과를 살펴볼 때, 학생의 자발적 정보 탐색에만 의존하는 것도 바람직하지 않다. 교사의 한정된 지식 영역을 뛰어넘고, 학생에게 길잡이 역할을 할 수 있는 교과연계 진로교육 콘텐츠의 마련이 시급하다고 볼 수 있다. 이상의 응답 및 분석 결과를 종합하면 다음과 같은 학생들의 수요가 도출된다.

- 진로에 대한 정보가 부족하여 진로 탐색 활동을 하거나, 전공을 선택하는 데 어려움을 느낀다.
- 사용 가능한 정보 채널이 학교 수업이나 포털 사이트 검색 등에 한정되어 있어 탐색 활동의 범위가 좁다.
- 진로 탐색 활동을 소수의 선생님께 의존하거나, 혼자서 실시하고 있어 관심 있는 전공을 넓게 탐색하기 어렵다.
- 진로 탐색을 위해서 무엇을 어떻게 해야 할지 모른다.

3.2 진로탐색 시스템 개발 요건의 정의

3.1에서 실시한 교사 대상 심층 인터뷰와 고등학생 대상 설문조사 결과를 바탕으로, 교사와 학생의 현장 요구를 모두 충족할 수 있는 교과연계 진로 콘텐츠 개발 및 추천 시스템의 요건은 다음 세 가지로 정의된다.

요건 1. 교육과정과 연계된 진로교육 콘텐츠가 데이터 기반 인공지능 기술을 활용하여 광범위하게 개발 및 추천되어야 한다. 이러한 콘텐츠 개발과 추천 기술을 통해, 학생에게 명확한 피드백을 제공할 수 있는 교과연계 진로교육 콘텐츠를 교과 교사가 이용하게 함으로써, 제도권 교육 안에서 맞춤형 진로교육이 원활하게 실시될 수 있다.

요건 2. 콘텐츠 개발 소스를 다각화하는 동시에 개발된 콘텐츠에 대한 정확성을 확보해야 한다. 다양한 분야의 전공자가 진로교육 콘텐츠 개발에 참여함으로써, 소수 교사가 짊어진 콘텐츠 개발의 부담을 분산시킬 수 있고, 동시에 다양하면서도 정교한 진로교육 콘텐츠를 개발할 수 있다.

요건 3. 시·공간 제약을 극복함과 동시에 콘텐츠의 균질성을 확보해야 한다. 한 명의 교사가 제한된 시·공간에서 다수의 학생을 상대해야 하는 기존의 교수·학습 시스템은, 맞춤형 진로교육을 효율적으로 실현하기에 한계가 있다. 따라서 진로교육 콘텐츠 개발자(전공자)와 이용자(학생 혹은 교사) 모두 온라인 환경에서 시·공간의 제약 없이 콘텐츠를 공급 및 이용할 수 있어야 하며, 이때 교육의 질적 수준을 보장할 수 있는 콘텐츠의 균질성이 확보되어야 한다.

4. 교과연계 진로탐색을 위한 데이터 기반 인공지능 시스템

본 장에서는 3장에서 정의한 시스템 개발 요건을 바탕으로, 1) 교과연계 진로교육 데이터 확보를 위한 데이터 정의와, 2) 확보된 데이터로부터 진로탐색 활동을 돕는 콘텐츠 개발 및 추천 인공지능 모델을 소개한다.

4.1 교과연계 진로교육 데이터 정의와 수집 방법

교육과정과 연계된 진로탐색 콘텐츠(요건 1)를 위한 데이터를 정의한다. 다음으로 콘텐츠의 정확성, 다각화와 균질성 확보(요건 2, 3)가 가능한 데이터 수집 방법을 소개한다.

데이터 정의 고교 교육과정 기반 진로탐색 콘텐츠 개발에 있어 핵심은, 1) 고등학생의 이수 교과(선택교과) 및 단원, 2) 고교 교과 과정 이후 수학하고자 하는 세부 전공, 그리고 3) 교과 및 전공과 연계된 탐구 주제라는 3가지 항목의 유기적인 연결이다.

데이터 내 교과 분류 첫째로, 고등학생이 이수 가능한 단원은 다음과 같이 [교과 > 교과 내 과목 > 과목 내 단원 > 단원 내 학습요소]로 분류할 수 있다. 분류 기준은 2015 개정 교육과정을 따랐으며[25], 본 연구는 3단계인 교과, 교과 내 과목, 과목 내 단원으로 연구 범위를 제한하였다. 단원 내 학습요소를 포함하는 4단계 분석은 추후 연구로 남긴다.

- **교과:** 고등학교 주요교과의 대분류로, 국어과, 수학과, 사회과, 과학과 4개로 분류한다.
- **교과 내 과목:** 각 교과에 속한 과목으로 전체 45개(국어과 8개, 수학과 10개, 과학과 13개, 사회과 14개)로 구성된다. 예컨대 과학과 내에는 [통합과학, 과학탐구실험, 물리학 I, 물리학 II, 화학 I, 화학 II, 생명과학 I, 생명과학 II, 지구과학 I, 지구과학 II, 과학사, 생활과 과학, 융합과학]와 같이 총 13개의 과목이 존재한다.
- **과목 내 단원:** 각 과목에 속한 단원으로, 전체 211개, 과목당 평균적으로 4.69개의 단원이 존재한다. 예컨대 과학과 내 융합과학 과목에는 [우주의 기원과 진화, 태양계와 지구, 생명의 진화, 정보통신과 신소재, 인류의 건강과 과학기술, 에너지와 환경] 총 6개의 단원이 존재한다.

데이터 내 전공 분류 두 번째로, 고등학생이 수학 가능한 전공은 다음과 같이 [대분류 > 중분류 > 전공(소계열)]로 분류할 수 있다. 분류는 대학알리미의 표준분류체계를 참조하되[26], 표준분류체계 내 미비학과 혹은 중복학과의 경우 자체적으로 추가 혹은 삭제하여 정비하였다.

- **대분류:** 전체 전공을 아우르는 분류 기준으로, 인문·사회, 자연과학, 공학, 의학, 교육 5개로 분류한다.
- **중분류:** 전공 등을 주제별로 아우르는 분류 기준으로 전체 24개(인문·사회 6개, 자연과학 7개, 공학 6개, 의학 1개, 교육 4개)로 구성된다. 공학을 예시로 들면 [전기·전자·컴퓨터, 건설, 산업·안전, 재료, 기계, 화공·고분자·에너지]와 같이 6개로 분류한다.
- **전공(소계열):** 중분류에 속한 전공 기준으로, 전체 138개, 중분류당 평균 5.75개의 전공이 존재한다. 공학 내 중분류 중 하나인 전기·전자·컴퓨터를 예시로 들면 [전기공학, 전자공학, 제어계측공학, 광학공학, 의공학, 응용소프트웨어공학, 전산학·컴퓨터공학, 정보·통신공학]와 같이 총 8개의 전공이 존재한다.

데이터 종합 앞서 정의한 데이터 내 교과 분류와 전공 분류를 바탕으로, 고교 교육과정과 수학 가능한 전공을 아우를 수 있는 진로 탐구 콘텐츠는 원리, 설명, 교과, 전공으로 구성되며, 그 예는 아래와 같다.

- **원리:** 진로 탐구 콘텐츠의 주제명
(예) 낙수효과
- **설명:** 원리에 대한 설명
(예) 고소득층의 소득이 증가하면 투자와 소비가 활성화되어 자연스럽게 저소득층의 소득도 증가한다는 이론이다. 분배보다는 성장에 중점을 둔 경제 이론이며, 실제 사회에서 처음 정식 활용된 사례로는 미국 40대 대통령 로널드 레이건의 고소득층 지지 정책이 있다.
- **교과:** [교과] - [과목] - [단원]
(예) 사회과 - 경제 - 경제생활과 경제 문제
사회과 - 사회문화 - 사회 계층과 불평등
수학과 - 경제 수학 - 함수와 경제
- **전공:** [대계열] - [중계열] - [소계열]
(예) 인문·사회 - 경영·경제 - 경제학
인문·사회 - 사회과학 - 사회학
자연과학 - 수학·물리·천문·지구 - 통계학

데이터 수집 및 검수 방법 앞서 정의한 시스템 개발 요건에 따라, 진로 탐색 콘텐츠 개발의 소스를 다각화하고, 정확성과 균질성을 확보하기 위하여 데이터는 다음과 같은 과정에 따라 수집 및 검수 되었다.

- [1단계] 학과별 전공자가 진로 탐구주제 생성
- [2단계] 학과별 전공자가 생성된 진로 탐구주제 검수
- [3단계] 검수된 탐구주제 교과/단원 태깅 초벌 작업
- [4단계] 데이터 학습을 통해 검수 및 태깅 작업 자동화

4.2 교과연계 진로교육 콘텐츠 개발, 추천 인공지능 모델

4.1장에서 정의한 데이터를 바탕으로 각 진로 탐구 주제에 따른 교육과정과 세부전공을 정의할 수 있다. 하지만 3.1에서 조사한 결과처럼, 진로·진학 지도 경험이 있는 국내 고등학교 교사라도 전체 단위 211개와 전체 계열(전공) 138개를 모두 이해한 다음 학생의 관심사에 따른 탐구 주제를 도출하는 것은 어렵다. 또한, 3.1에서 조사하고 3.2에서 도출한 결과처럼, 학생들은 사전에 본인이 이수하고자 하는 선택교과와, 수학하고자 하는 전공 조합에 대한 사전 조사를 바탕으로, 본인의 탐구 주제를 선택하는 경우가 드물다. 따라서 이수가 끝났거나 이수 예정인 선택교과 중, 지적 호기심을 바탕으로 탐구하고자 하는 선택교과 내 단원을 선택하면, 해당 단원과 관련된 전공을 추천해주는 등의 데이터 기반 인공지능 모델이 필요하다. 본 장은 3장에서 조사하고 도출한 결과와 4.1에서 정의한 데이터를 바탕으로, 인공지능 기술을 바탕으로 해결 가능한 문제를 정의하고, 이를 해결하는 방법을 기술한다.

문제정의; Problem Definition 3.1에서 조사하고 3.2에서 도출한 결과와 같이, 교사 입장에서 각 탐구 주제에 대해 전체 단위 211개와 전체 계열 138개를 모두 이해한 다음 탐구 주제를 도출하는 것은 어렵다. 또한, 학생의 진로 탐구를 돕기 위해서는, 학생 본인이 이수한 교과를 바탕으로 세부전공을 추천하거나, 관심 있는 세부전공을 뒷받침하는 선택교과를 추천해주는 모델이 필요하다. 따라서 본 연구에서 데이터 기반 인공지능 모델로서 해결하고자 하는 문제는, 고교 과정에서 이수할 수 있는 선택교과와, 고교 과정을 마치고 수학 가능한 세부전공에 대한 명확한 정보 제공이라고 할 수 있다.

다시 말해, 진로탐구 콘텐츠 개발의 효율성을 높이고, 학생의 선택교과 또는 세부전공 선택을 도울 뿐만 아니라, 교사의 진로·진학 지도를 돕기 위해서는, 선택교과를 바탕으로 이루어지는 세부 전공의 추천과 세부전공을 바탕으로 이루어지는 선택교과의 추천이 가능한 인공지능 모델이 필요하다.

인공지능 학습 목표; Learning Objective 각 진로탐구 주제 T_i 는 선택교과의 집합 X_i 과, 세부전공의 집합 X_m 으로 구성되어있다. 앞의 예시를 바탕으로 보면, 진로탐구 주제인 낙수 효과 T_1 은, 선택교과의 집합인 X_1 ={사회과, 경제, 경제생활과 경제 문제}, (사회과, 사회문화, 사회 계층과 불평등), (수학과, 경제 수학, 함수와 경제))와, 세부전공의 집합인 X_m ={((인문·사회, 경영·경제, 경제학), (인문·사회, 사회과학, 사회학), (자연과학, 수학·물리·천문·지구, 통계학))}으로 구성된다.

X_i 또는 X_m 와 같은 셋을 인코딩하는 미분가능한 함수를 f 로 정의하고, f 의 결과를 바탕으로 셋을 디코딩하는 미분가능한 함수를 g 로 정의하자. 이때, $g(f(X_i))$ 의 결과를 Y_i 라고 하고, $g(f(X_m))$ 의 결과를 Y_m 라고 하면, 인공지능 모델의 학습 목표는 L 이라는 손실 함수와 함께 다음과 같이 정의된다:

$$\begin{aligned} \min L(Y_i - X_m), Y_i &= g(f(X_i)) \text{ or} \\ \min L(Y_m - X_i), Y_m &= g(f(X_m)). \end{aligned} \quad (1)$$

이는 선택교과의 집합을 입력으로 받아 세부전공의 집합을 추천하는 문제이며, 반대로 세부전공의 집합을 입력으로 받

아 선택교과의 집합을 추천하는 문제로 표현된다.

셋 기반의 인코딩; Set-based Encoding 집합의 원소를 포함하는 문제는 반드시 permutation invariant property 성질을 만족해야 한다. 이 의미는 주어진 집합의 원소 순서가 바뀔지라도, target value의 값은 변하지 않아야 한다는 것이다. 이를 위해서 모든 permutation invariant 함수 NN 은 다음과 같이 표현된다[8]:

$$NN(x_1, \dots, x_n) = p(\text{pool}(\phi(x_1), \dots, \phi(x_n))). \quad (2)$$

이때 x_i 는 앞서 정의한 X_i 또는 X_m 집합 내의 원소들이며, pool 은 sum operation을 나타내고, p 와 ϕ 함수는 어떠한 continuous 함수가 가능하다. 또한 n 은 집합 내의 원소의 개수이다.

또한, 만약 모델의 encoder가 permutation equivalent layer들이 쌓여있는 형태라면, 모델은 permutation invariant 한 성질을 잃지 않는다. permutation equivalent layer는 다음과 같이 정의된다:

$$f_i(x; x_1, \dots, x_n) = \sigma_i(\lambda x + \gamma \text{pool}(x_1, \dots, x_n)). \quad (3)$$

이때, pool 은 pooling 함수이며, λ 와 γ 는 학습 가능한 scalar variables이다. 또한 σ 는 non-linear activation function을 나타낸다. 본 연구에서는 위와 같은 permutation equivalent layer들을 쌓아 X_i , X_m 셋을 인코딩할 것이다. 예를 들어, 인코딩 함수 f 을 $f = f_3(f_2(f_1(\cdot)))$ 와 같이 세 개의 permutation equivalent layer를 쌓아 표현할 수 있다. 본 연구에서 사용하는 집합에 대한 인코딩 함수의 아웃풋을 다음과 같이 정의하자:

$$z_m = f(X_m), \quad z_i = f(X_i). \quad (4)$$

셋 기반의 디코딩; Set-based Decoding 셋 기반 디코딩 함수의 목표는, 인코딩 함수 f 을 이용해서 표현 공간에 표현된 집합 X_i 또는 X_m 을 바탕으로, 집합 X_m 또는 X_i 를 각각 복구하는 것이다. 이는 선택교과 집합 X_i 를 입력으로 받아 세부전공 집합 X_m 을 출력으로 추천할 수 있으며, 반대로 세부전공 집합 X_m 을 입력으로 받아 선택교과 집합 X_i 를 추천할 수 있게 한다. 이를 위해 우리는 셋 기반의 인코딩 함수 f 의 아웃풋인 z 을 입력으로 받은 다음, MLP와 같은 continuous function을 사용한 디코딩 함수인 g 을 통해 추천하고자 하는 집합을 예측한다. 이는 다음과 같이 정의된다:

$$SM(g(z)) = SM(Y), \quad Y = y_1, \dots, y_N. \quad (5)$$

이때, SM 은 softmax 함수, N 은 예측하고자 하는 집합 Y 가 가질 수 있는 전체 원소의 개수를 의미한다. 예를 들어, 예측하고자 하는 집합을 X_m 라고 할 때, 예측 가능한 전공의 전체 개수는 4.1에서 정의한 것과 같이 138개이다; $N=138$. 또한 softmax 함수는 집합의 원소 단위로 적용되며, 다음과 같이 정의된다:

$$SM(y_i) = \frac{\exp(y_i)}{\sum_{i=1}^N \exp(y_i)}. \quad (6)$$

결과적으로, Y 의 원소가 가지는 값을 바탕으로, 추천하고자 하는 집합 셋의 원소들을 예측할 수 있다.

손실함수; Loss Function 본 인공지능 모델의 목표는 집합 X_i 또는 X_m 가 주어졌을 때, 인코딩 함수와 디코딩 함수인 f 와 g 을 거쳐, X_m 또는 X_i 을 예측하는 것이다. 이러한 예측 모델을 구성하는 함수인 f 와 g 을 학습하기 위해, 본 연구는 손실함수 L 로서 cross-entropy loss를 사용한다. cross-entropy loss는 negative log likelihood를 최소화하는 방향으로 학습이 되며, 다음과 같이 정의된다:

$$\begin{aligned} \min \sum_{i=1}^n -X_{m_i} \log(Y_i), \quad Y_i = g(f(X_i)) \text{ or} \\ \min \sum_{i=1}^n -X_i \log(Y_m), \quad Y_m = g(f(X_m)). \end{aligned} \quad (7)$$

5. 실험

본 장에서는 데이터 구성과 실험 구성 방식을 설명한 다음, 교과연계 진로탐색 시스템을 위한 인공지능 모델의 실험 결과와 현장 반응에 대해 설명한다. 인공지능 모델의 성능 분석을 위한 실험은 다음과 같은 2가지 방법으로 진행된다. 첫 번째(A. Soft Problem)는 1) 교과연계 진로탐구 주제, 2) 고교 교과 분류 내 과목, 3) 전공(소계열)의 유기적 관계를 인공지능 모델이 추론할 수 있는지 살펴본다. 두 번째(B. Hard Problem)는 1) 교과연계 진로탐구 주제, 2) 고교 교과 과목 내 단원, 3) 전공(소계열)의 유기적 관계를 인공지능 모델이 추론할 수 있는지 살펴본다. 첫 번째 문제는 교과 내 과목이라는 고교 교과 분류 중 2단계 분류를 사용한 것이고, 두 번째 문제는 과목 내 단원이라는 고교 교과 분류 중 3단계 분류를 사용한 것이다.

Table 1. Dataset Statistics for Soft and Hard Problems

	The number of contents for career exploration	The average number of lectures per content	The average number of majors per content
A. Soft Problem	1,027	6.53	6.71
B. Hard Problem	871	6.78	5.55

5.1 데이터 구성

진로탐색 콘텐츠를 위한 고교 교과 분류는 [교과 > 교과 내 과목 > 과목 내 단원]과 같이 3단계로 분류된다. 교과는 4개, 교과 내 과목은 45개, 과목 내 단원은 211개의 항목으로 구성된다. 또한, 진로탐색 콘텐츠를 위한 전공 분류는 [대분류 > 중분류 > 전공(소계열)]과 같이 3단계로 분류된다. 대분류는 5개, 중분류는 24개, 전공은 138개의 항목으로 구성된다. A, B 문제를 위한 데이터는 4.1에서 정의한 데이터 수집 및 검수 방법을 따르며, 각 데이터에 대한 통계는 Table 1에서 기술한다.

A. Soft Problem A 문제를 위한 데이터는 1) 교과연계 진로탐구 주제, 2) 고교 교과 분류 내 과목 45개, 3) 전공(소계열) 138개의 조합으로 구성된다. 해당 문제를 위한 1), 2), 3)의 조합으로 1,027개의 데이터를 수집했으며, 교과연계 진로탐구 주제당 평균적으로 6.53개의 고교 교과 내 과목이 있고, 6.71개의 전공(소계열)이 존재한다.

B. Hard Problem B 문제를 위한 데이터는 1) 교과연계 진로탐구 주제, 2) 고교 교과 분류 내 과목 211개, 3) 전공(소계열) 138개의 조합으로 구성된다. 해당 문제를 위한 1), 2), 3)의 조합으로 871개의 데이터를 수집했으며, 교과연계 진로탐구 주제당 평균적으로 6.78개의 고교 교과 분류 내 단원이 있고, 5.55개의 전공(소계열)이 존재한다.

5.2 실험 정의

본 연구에서는 앞서 소개한 것처럼, A. Soft Problem(교과연계 진로탐구 주제, 고교 교과 분류 내 과목, 전공 3가지 콘텐츠의 유기적 관계)과 B. Hard Problem(교과연계 진로탐구 주제, 고교 교과 과목 내 단원, 전공 3가지 콘텐츠의 유기적 관계) 문제를 인공지능 모델이 이해하고, 이에 따른 콘텐츠 추천이 가능한지 살펴보기 위해 다음과 같은 두 가지 문제 각각에 대한 실험을 정의하였다.

A. Soft Problem A 문제를 인공지능 모델이 이해한 다음, 그에 따른 진로 탐구 콘텐츠(선택교과 또는 세부전공) 추천이 가능한지 알아보기 위해, 첫 번째로 (**A. Task #1. Major Recommendation**), 고교 교과 분류 내 과목이 주어졌을 때, 전공(소계열) 추천이 가능한지 살펴본다. 두 번째로 (**A. Task #2. Lecture Recommendation**), 전공(소계열)이 주어졌을 때 고교 교과 분류 내 과목 추천이 가능한지 살펴본다.

B. Hard Problem 앞서 정의한 A 문제에 대한 두 가지 Task #1, #2와 유사하게, B 문제를 인공지능 모델이 이해한 다음, 그에 따른 진로 탐구 콘텐츠(선택교과 또는 세부전공) 추천이 가능한지 알아보기 위해, 첫 번째로 (**B. Task #1. Major Recommendation**), 고교 교과 과목 내 단원이 주어졌을 때, 전공(소계열) 추천이 가능한지 살펴본다. 두 번째로 (**B. Task #2. Lecture Recommendation**), 전공(소계열)이 주어졌을 때 고교 교과 과목 내 단원 추천이 가능한지 살펴본다.

A 문제와 B 문제의 차이점은, A 문제는 고교 교과 분류 내 과목을 바탕으로 진로탐구 주제를 설정하고, B 문제는 과목 내에 있는 단원을 바탕으로 진로탐구 주제를 설정하는 것이다. 따라서 B 문제의 경우, 과목 내에 있는 단원을 바탕으로 탐구 주제를 이해하기 위해서는, A 문제에 비해 더 많은 변수 간의 상호관계를 살펴봐야 함으로 A 문제에 비해 사람이 더 해결하기 어려운 주제이다.

5.3 실험 구현

앞서 제안한 각각의 문제에 대한 데이터는 약 1,000개 정도로 답러닝이 적용되는 이미지[27] 또는 자연어[28] 문제들

에 비해 상대적으로 적은 편이다. 따라서 보다 정확한 실험 환경 구현 및 평가를 위해 K-Fold Cross Validation을 사용한다. K-Fold Cross Validation에서, K 값은 10으로 두었으며, 학습 데이터의 10%를 평가 데이터로 사용했다. 모델을 최적화하기 위해 Adam Optimizer[29]를 사용하였고, 각 실험 Fold마다 300번의 iteration을 거치며 학습을 진행했다. Learning rate 값은 0.01, Grad norm 값은 1.0, 각 layer마다 dropout 비율은 0.3, 모델 각 layer의 hidden size는 64로 구성했다. 인코더와 디코더로는 각각 2개의 feed-forward layer를 사용하였으며, 인코더의 pooling 함수로는 sum operation을 사용했다. 마지막으로, 모델 성능의 정량적 평가는, 모델이 예측한 N개의 값 중 Top-n개의 항목이, 실제로 예측 해야 할 n개의 항목과 일치하는지를 바탕으로 두는 accuracy 기반 평가 방식을 사용했다.

5.4 정량적 실험 결과

셋 기반의 인코딩-디코딩 구조를 바탕으로 설계한, 진로 콘텐츠 추천 인공지능 모델의 성능이 Table 2와 같이 random guess 성능을 압도적으로 뛰어넘은 것을 확인하였다. 더 나아가, Task #2. Lecture Recommendation의 경우 Problem A가 가지는 45개의 정답 후보에 비해, Problem B가 가지는 정답 후보 211개가 압도적으로 많아 제안한 모델의 성능이 일부 감소하는 것을 보인다. 하지만 Task #1. Major Recommendation에서 알 수 있듯이, 45개의 교과 정보만을 사용하는 것보다, 세분화된 211개의 단원 정보를 사용하는 것이 세부전공 추천 성능을 높이는 것을 확인하였다.

Table 2. 10-fold Cross Validation Results for Random Guess Baseline and our Models for Soft and Hard Problems with Accuracy (Standard Deviation) as an Evaluation Metric

	A. Soft Problem		B. Hard Problem	
	Task #1. Major *rd.	Task #2. Lecture rd.	Task #1. Major rd.	Task #2. Lecture rd.
Random Guess	0.049	0.145	0.040	0.032
Ours	0.567 (0.022)	0.590 (0.023)	0.583 (0.027)	0.398 (0.029)

*rd.=recommendation

5.5 정성적 실험 결과

아래의 (a), (b), (c), (d)는 수행한 각각의 실험에 대해 제안한 모델이 보인 정성적 실험 결과를 나타낸다. 해당 실험 결과처럼 실제 값과 유사하거나, 실제 값보다 더 탐구 주제에 어울리는 값을 모델이 예측하는 것을 확인할 수 있다.

(a) A. Soft Problem, Task #1. 전공 추천

[원리] 네덜란드 병

[설명] 네덜란드 병은 자원이 풍부한 나라가 자원을 수출하여 경제적 호황을 누리더라도, 산업 경쟁력을 갖추는 것을 등한시하다가 결국 경제 침체를 맞는다는 내용이다. 1959년 네덜란드가 유전을 발견한 후 수출을 통해 막대한 수입을 얻었지만, 곧 물가와 통화 가치가 상승하고 국내 제조업이 쇠퇴함으로써 경제 침체를 맞은

데에서 유래되었다. 자원의 저주라고도 불린다.

[교과] 경제, 통합사회, 세계지리

[실제 전공] 사회학, 정치외교학, 국제학

[예측 전공(예측 수치)]

경제학(0.953), 경영학(0.019), 사회학(0.005)

(b) A. Soft Problem, Task #2. 교과 추천

[원리] 레몬 시장 이론

[설명] 레몬 시장 이론은 미국 경제학자 애컬로프가 1970년에 발표한 이론으로, 판매자와 구매자 간 정보의 비대칭성으로 인해 시장에 존재하는 재화나 서비스의 품질이 낮아진다는 이론이다. 애컬로프는 중고차 시장에서 판매자가 구매자에게 정보를 투명하게 공개하지 않아 구매자들이 어쩔 수 없이 질 낮은 중고차를 구매하게 되는 상황에서 이 이론을 떠올렸다. 품질이 낮은 중고차를 레몬이라고 부른 데에서 레몬 시장이라는 단어가 유래되었다.

[전공] 경제학, 정치외교학, 행정학

[실제 교과] 정치와 법, 경제, 사회문화

[예측 교과(예측 수치)]

경제(0.482), 통합사회(0.279), 사회문화(0.108)

(c) B. Hard Problem, Task #1. 전공 추천

[원리] 성악설

[설명] 성악설이란 인간의 본성이 악하다고 주장하는 학설이다. 대표적인 유가 철학자인 순자는 성악설을 지지했다. 이와 대비되는 개념으로는 성선설, 성무선악설이 있다. 유가의 또 다른 대표적 철학자인 맹자는 순자와 달리 인간의 본성이 선하다는 의미의 성선설을 주장했으며, 또 다른 철학자 고자는 인간의 본성은 선하지도 악하지도 않다고 주장하는 성무선악설을 지지했다.

[교과] 인간과 윤리사상, 인간·사회·환경과 행복, 민주주의와 헌법

[실제 전공] 종교학, 인류학, 철학·윤리학

[예측 전공(예측 수치)]

철학·윤리학(0.709), 사회학(0.212), 법학(0.029)

(d) B. Hard Problem, Task #2. 교과 추천

[원리] 불쾌한 골짜기

[설명] 일본 로봇공학자 모리 마사히로가 1970년에 처음 발표한 이론으로, 로봇이 인간과 비슷한 모습을 할수록 인간이 로봇에게 느끼는 호감도가 증가하지만, 그 비슷한 정도가 특정 수준에 다다르면 오히려 강한 혐오감과 불쾌감을 느낀다는 내용이다. 하지만 그 수준을 넘어 인간과 완전히 동일한 수준에 이르면 다시 호감도가 증가한다.

[전공] 철학·윤리학, 심리학, 사회학

[실제 교과] 인간·사회·환경과 행복, 타인과의 관계, 거주 공간의 변화와 지역 개발

[예측 교과(예측 수치)]

생활공간과 사회(0.179), 과학과 현대 사회(0.092), 인간·사회·환경과 행복(0.088)

5.6 현장 반응

본 연구에서 개발한 추천 시스템의 활용성을 검증하고자, 교육 현장의 학생 300명을 대상으로 콘텐츠를 이용하도록 한 후 설문 조사 및 인터뷰를 실시하였다. 이때 교육 정보 접근성이 비교적 높은 수도권과, 접근성이 다소 낮은 지방권 학생 간에 만족도 차이를 보일 수 있다고 판단하여, 전국 고등학생의 현장 반응을 두루 검증하고자 온라인 학습커뮤니티에서 대상자를 지원 받았다.

전국 고등학생 300명에게 본 연구의 추천 시스템을 활용한 진로탐색 콘텐츠를 제공한 후, 제공받은 콘텐츠 대하여 1) 희망 전공과의 관련성, 2) 과목과의 관련성, 3) 단원과의 관련성, 4) 난이도의 적절성, 5) 재이용 의사를 5점 만점의 리커트 척도로 평가하도록 하였다. 설문 결과는 아래와 같다.

- 1) 희망 전공과의 관련성: 4.19 / 5.0
- 2) 과목과의 관련성: 4.60 / 5.0
- 3) 단원과의 관련성: 4.46 / 5.0
- 4) 난이도의 적절성: 3.55 / 5.0
- 5) 재이용 의사: 4.67 / 5.0

설문 결과 희망 전공과의 관련성, 과목과의 관련성, 단원과의 관련성, 재이용 의사 모두 4점 이상의 높은 점수를 받았다. 특히 재이용 의사에서 가장 높은 점수를 받은 것은 본 연구에서 개발한 추천 시스템의 현장 활용 가능성이 매우 높음을 시사한다. 한편 전반적으로 지방 고등학교 학생의 만족도가 특히 높게 나타났는데, 이는 본 연구에서 개발한 추천 시스템이 수도권과 지방권의 정보 격차를 해소하는 데에도 기여할 수 있음을 암시한다.

다만 난이도의 적절성의 경우 다른 항목에 비해 다소 낮은 점수를 보였는데, 그 원인을 파악하고자 추가적으로 인터뷰를 실시하였다. 그 결과 ‘개념이 생소해서 다소 어려웠다’, ‘관련 보조자료가 있으면 좀 더 도움이 될 것 같다’, ‘(과학 과목의 경우) 문과 학생이라 조금 어려웠다’ 등의 의견이 두드러졌다. 그런데 이는 학생들이 교과서나 인터넷 검색만으로 쉽게 접하기 어려운 진로 탐색 콘텐츠를 제공함에 따른 결과로서, 오히려 깊이있고 실질적인 맞춤형 진로교육을 지향하는 본 추천 시스템의 본질을 보여준다고 해석할 수 있다. 한편 이러한 난이도의 문제는 교사의 보충 설명이나 보조자료 제공 등을 통해 보완할 수 있을 것으로 기대한다.

6. 결 론

본 연구는 진로교육 중심의 교육 패러다임 변화에 따라, 학생들의 자율적인 진로탐색을 지향하는 자유학기제와 고교학점제의 정책적 목표를 바탕으로, 교과연계 진로교육을 향한 에듀테크 기술이 추구해야 하는 요구조건 세 가지를 정의했다. 그 다음, 데이터 기반 기술을 통해 진로 탐구 주제와 고교 과목 및 수학 가능한 세부 전공을 아우르는 데이터 시스템 및 인공지능 추천 모델을 제안했다. 마지막으로 실험을 통해, 제안한 인공지능 모델이 진로교육 콘텐츠 추천에 있어 만족할 만한 성능을 보이는 것을 확인하였으며, 사용자 측면에 있어 교육 현

장에서의 실제 적용 결과가 만족스럽다는 것도 확인하였다.

진로 분야의 교육적 측면에서 본 연구를 활용하면, 교사 측면에서 존재하는 모든 학과와 직업 정보를 알지 않아도 학생 개개인의 흥미와 관심사에 따른 진로지도가 가능해진다. 또한 학생 측면에서 교과 관련 진로 정보가 부족하여 발생했던 제한적 탐색 범위를 본 데이터 기반 인공지능 모델을 통해 넓힐 수 있을 것이라 기대한다.

모델 구조적 측면에서 본 연구의 인공지능 모델은 셋 기반의 인코딩-디코딩 모델로, 진로 탐구 주제를 바탕으로 파생되는 고교 선택교과 및 수학 가능한 세부 전공을 벡터 공간에 표현하였다. 하지만 이뿐만 아니라 진로 탐구 주제 텍스트 그 자체도 자연어 기술을 이용하여 인코딩한 다음, 고교 선택교과 또는 수학 가능한 세부 전공을 추천함에 있어 보조 정보로 활용할 수 있다. 이에 대한 연구는 추후 연구로 남긴다.

References

- [1] J. Y. Choi et al., "This is the 4th industrial revolution," *Maeil Business News*, 2017.
- [2] Ministry of Education, "A Plan for the Implementation of the Free Semester System in Middle School," 2016.
- [3] Samcheok Office of Education, The Concept of Free School Year System [Internet], <http://www.kwsche.go.kr/sub/info.do?m=020601>.
- [4] Korean Educational Development Institute, Introduction of High School Credit System [Internet], <https://www.hscredit.kr/common/menu/html/900900001/detailPotal.do>.
- [5] N. Kim, H. Lee, Y. Lee, Y. Jeong, and Y. Cho, "Analysis of experiences and effects of high school teachers' subjects-related career education classes using grounded theory," *The Study of Career Education*, Vol.31, No.2, pp.27-50, 2018.
- [6] E. Hyeon, "The Effect of Free Semester Creative Career Education Program on Media Convergence Design," NRF KRM(Korean Research Memory), 2015.
- [7] J. Lee, M. Park, M. So, and S. Ahn, "Corona19 and Education: Focusing on the Life and Perception of School Members," *Gyeonggi Institute of Education*, pp.162-168, 2020.
- [8] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczós, R. Salakhutdinov, and A. J. Smola "Deep Sets," in *Proceedings of Advances in Neural Information Processing Systems*, pp.3391-3401, 2017.
- [9] Free-Semester Website [Internet], www.ggoomggi.go.kr.
- [10] J. Kim, "A Study on Changes to Students Academic Records according to Their Satisfaction with the Free Semester System before and after Its Implementation (2015-2017)," *The Journal of Learner-Centered Curriculum and Instruction*, Vol.18, No.18, pp.293-316, 2018.
- [11] S. Lee and S. Baek, "Exploring the Possibility of Expanding the Choice of High School Credit System," *The Journal of Korean Teacher Education*, Vol.36, No.2, pp.49-73, 2019.

- [12] G. Park, Y. Joo, and I. Lee, "An analysis on the role of school members for stable settlement of the High School Credit System," *Journal of Research in Education*, Vol.32, No.4, pp.87-113, 2019.
- [13] Y. Kim and A. Choi, "Teacher Perception and Practice on Free Semester Science Assessment," *Journal of the Korean Association for Science Education*, Vol.39, No.1, pp.143-160, 2019.
- [14] J. Seo and S. Shin, "A Case Study on the Effectiveness of Major-friendly Contents in Software Education for the Non-majors," *Journal of Digital Convergence*, Vol.18 No.5, pp.55-63, 2020.
- [15] H. Seo and S. Kyun, "An analysis of research published in Korean Journal with the theme of 'education & AI', using semantic network analysis," *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, Vol.9 No.11, pp.209-217, 2019.
- [16] W. Shin and D. Shin, "A Study on the Application of Artificial Intelligence in Elementary Science Education," *Elementary Science Education*, Vol.39, No1, pp.117-132, 2020.
- [17] B. Shi, S. Bai, Z. Zhou, and X. Bai, "DeepPano: Deep Panoramic Representation for 3-D Shape Recognition," *IEEE Signal Processing Letters*, Vol.22, No.12, pp.2339-2343, 2015.
- [18] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," *IEEE Conference on Computer Vision and Pattern Recognition*, pp.77-85, 2017.
- [19] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. J. Guibas, "Learning Representations and Generative Models for 3D Point Clouds," *Proceedings of the 35th International Conference on Machine Learning*, pp.40-49, 2018.
- [20] L. Yi, W. Zhao, H. Wang, M. Sung, and L. J. Guibas, "GSPN: Generative Shape Proposal Network for 3D Instance Segmentation in Point Cloud," *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp.3947-3956.
- [21] Y. Lee and S. Choi, "Gradient-Based Meta-Learning with Learned Layerwise Metric and Subspace," *Proceedings of the 35th International Conference on Machine Learning*, pp.2933-2942, 2018.
- [22] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical Networks for Few-shot Learning," *Advances in Neural Information Processing Systems*, pp.4077-4087, 2017.
- [23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is All you Need," *Advances in Neural Information Processing Systems*, pp.5998-6008, 2017.
- [24] J. Lee, Y. Lee, J. Kim, A. R. Kosiorek, S. Choi, and Y. W. Teh, "Set Transformer: A Framework for Attention-based Permutation-Invariant Neural Networks," *Proceedings of the 35th International Conference on Machine Learning*, pp.3744-3753, 2019.
- [25] National Curriculum Information Center, 2015 Curriculum [Internet], <https://ncic.re.kr/>.
- [26] Higher Education in Korea, Standard Classification System [Internet], <https://www.academyinfo.go.kr/mjrinfo/mjrinfo0460/doInit.do>.
- [27] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F. Li, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, Vol.115 No.3, pp.211-252, 2015.
- [28] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp.4171-4186, 2019.
- [29] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations*, 2015.



백진헌

<https://orcid.org/0000-0002-9367-560X>
 e-mail : jinheon.baek@kaist.ac.kr
 2020년 고려대학교 컴퓨터학과(학사)
 2020년 ~ 현 재 한국과학기술원
 인공지능대학원 석사과정
 관심분야 : Machine Learning, Graph
 Neural Network



김하연

<https://orcid.org/0000-0003-3495-8236>
 e-mail : khy3221@snu.ac.kr
 2016년 서울대학교 국어교육과(학사)
 2020년 서울대학교 국어교육과(석사)
 2020년 ~ 현 재 서울대학교 국어교육과
 박사과정

2020년 ~ 현 재 (주)잡쇼퍼 교육본부장

관심분야 : Career Education, Writing, Prose Education



권기원

<https://orcid.org/0000-0002-0415-6550>
 e-mail : kiwoun203@knou.ac.kr
 2018년 고려대학교 경영학과(학사)
 2017년 ~ 현 재 (주)잡쇼퍼 대표이사
 2017년 ~ 현 재 고려대학교 SW중심대학
 자문위원

2019년 ~ 현 재 한국방송통신대학교 이터닝학과 석사과정

관심분야 : Artificial Intelligence, Career Education