

Design and Implementation of YouTube-based Educational Video Recommendation System

Young Kook Kim*, Myung Ho Kim*

*Student, Dept. of Software, Soongsil University, Seoul, Korea

*Professor, Dept. of Software, Soongsil University, Seoul, Korea

[Abstract]

As of 2020, about 500 hours of videos are uploaded to YouTube, a representative online video platform, per minute. As the number of users acquiring information through various uploaded videos is increasing, online video platforms are making efforts to provide better recommendation services. The currently used recommendation service recommends videos to users based on the user's viewing history, which is not a good way to recommend videos that deal with specific purposes and interests, such as educational videos. The recent recommendation system utilizes not only the user's viewing history but also the content features of the item. In this paper, we extract the content features of educational video for educational video recommendation based on YouTube, design a recommendation system using it, and implement it as a web application. By examining the satisfaction of users, recommendation performance and convenience performance are shown as 85.36% and 87.80%.

▶ **Key words:** Machine Learning, Deep Learning, Speech Analysis, Recommendation System, System Implementation

[요 약]

2020년 기준 대표적인 온라인 동영상 플랫폼인 유튜브에는 1분에 약 500시간의 동영상이 업로드되고 있다. 이에 업로드된 다수의 다양한 동영상을 통해 정보를 획득하는 사용자의 수가 늘고 있어 온라인 동영상 플랫폼들은 더 나은 추천 서비스를 제공하기 위해 노력하고 있다. 현재 사용되고 있는 추천 서비스는 사용자의 시청 기록을 기반으로 사용자에게 동영상을 추천하는데 이는 교육용 동영상과 같이 특정 목적 및 관심사를 다루는 동영상 추천에 좋은 방법이 아니다. 최근 추천 시스템은 사용자의 시청 기록뿐만 아니라 아이템의 콘텐츠 특징을 함께 활용한다. 본 논문에서는 유튜브를 기반으로 교육용 동영상 추천을 위한 교육용 동영상의 콘텐츠 특징을 추출하고, 이를 활용하는 추천 시스템을 설계하여 웹 애플리케이션으로 구현한다. 사용자들의 만족도를 조사하여 추천 시스템의 추천 성능의 만족도 85.36%, 편의성 만족도 87.80%를 보인다.

▶ **주제어:** 머신러닝, 딥러닝, 음성 분석, 추천 시스템, 시스템 구현

-
- First Author: Young Kook Kim, Corresponding Author: Myung Ho Kim
 - *Young Kook Kim (1101978003@soongsil.ac.kr), Dept. of Software, Soongsil University
 - *Myung Ho Kim (kmh@ssu.ac.kr), Dept. of Software, Soongsil University
 - Received: 2022. 02. 17, Revised: 2022. 05. 09, Accepted: 2022. 05. 09.

I. Introduction

최근 코로나19의 영향으로 온라인 동영상 서비스(Over The Top; OTT)가 빠르게 성장하고 있다. 대표적인 온라인 동영상 플랫폼인 유튜브는 2020년 기준 다양한 장르의 동영상들이 1분에 약 500시간 정도 업로드되고 있다[1]. 이에 온라인 동영상 서비스 플랫폼들은 고객 유치를 위한 방안으로 사용자들에게 더 나은 추천 서비스를 제공하기 위한 노력을 하고 있다.

현재 온라인 동영상 서비스 플랫폼들이 주로 사용하고 있는 추천 서비스는 사용자의 시청 기록을 기반으로 동영상의 카테고리, 제목 및 동영상 채널, 조회 수, 좋아요와 같은 동영상의 기본적인 특징을 활용한다. 기존 추천 서비스는 여유 시간을 활용하는 킬링타임 측면에서 고객이 관심 있는 분야의 동영상 및 기존에 고객이 만족한 영상과 유사한 동영상을 추천해주면서 고객을 플랫폼에 머무르게 하는 역할을 잘 수행하고 있다.

Fig. 1의 SM C&C 광고사업 부분 설문 조사 플랫폼 Tillion Pro가 유튜브 이용 경험자 984명을 대상으로 유튜브 주 이용 목적을 조사한 결과에 따르면 전체 이용자의 25.9%가 새로운 정보 탐색 및 정보 습득을 목적으로 유튜브를 이용한다[2]. 하지만 새로운 정보 탐색 및 습득을 목적으로 하는 고객의 수는 증가하지만, 기존의 온라인 동영상 서비스 플랫폼의 추천 서비스는 특정 주제 및 교육을 위한 동영상 추천에 특화된 서비스로 볼 수 없다.

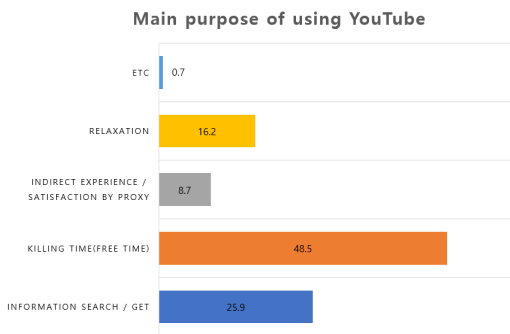


Fig. 1. Research on the Purpose of using YouTube

본 논문은 온라인 동영상 플랫폼 중 유튜브를 기반으로 새로운 정보 탐색 및 습득을 목적으로 하는 교육용 동영상 추천 시스템을 설계 및 구현한다. 최근 추천 시스템은 사용자와 아이템 사이의 관계뿐만 아니라 아이템의 콘텐츠 특징을 활용하여 더 나은 서비스를 제공하고 있다. 따라서 본 논문이 제안하는 추천 서비스는 기존 온라인 동영상 서

비스 플랫폼들의 추천 서비스가 사용하는 동영상의 제목, 조회 수 등의 기본적인 특징 외에 동영상에 가지는 교육용 콘텐츠 특징을 추출하여 추천 시스템에 적용한다.

본 논문은 다음과 같이 구성된다. 2장에서는 콘텐츠 특징을 활용한 기존의 추천 방법을 설명하고, 교육용 동영상 추천에 필요한 콘텐츠 특징을 파악한다. 3장에서는 교육용 동영상의 콘텐츠 특징을 추출하기 위한 모델과 추천 시스템을 설계한다. 4장에서는 설계한 추천 시스템을 요약하고, 5장에서 결론 및 미래 연구를 기술한다.

II. Preliminaries

1. Contents based Recommendation System

기존의 추천 시스템은 사용자들의 아이템에 대한 기호정보를 바탕으로 협업 필터링 방법에 적용하여 사용자의 취향을 분석한다. 협업 필터링은 사용자의 취향을 파악하는 합리적인 방법이지만, 사용자들의 정보가 많아 제대로 추천할 수 있을뿐더러 콜드 스타트 문제를 가지고 있다. 이런 한계를 극복하는 방법으로 콘텐츠 기반 필터링이 있다.

콘텐츠 기반 필터링은 아이템에 대한 분석을 통한 특징인 콘텐츠를 기반으로 추천하는 기법이다[3, 4]. 예를 들어 영화를 분석할 때, 장르, 감독, 배우 등의 정보를 콘텐츠 특징 또는 메타데이터라고 한다. Fig. 2는 콘텐츠 기반 필터링을 도식화한 것으로 ‘라이언 일병 구하기(Saving Private Ryan, 1998)’, ‘캐치 미 이프 유 캔(Catch Me If You Can, 2002)’을 좋아하는 사용자에게 같은 감독 및 주연 배우가 출연한 작품인 ‘더 포스트(The Post, 2017)’를 추천한다.

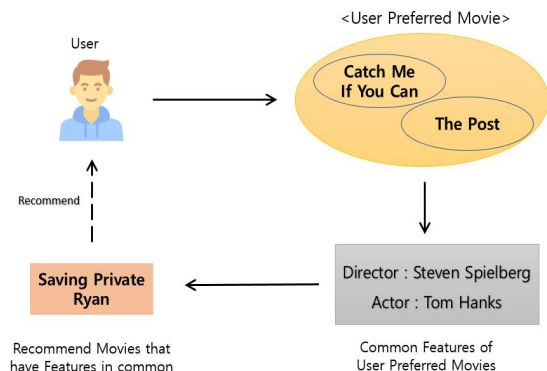


Fig. 2. Example of Content-based Filtering

콘텐츠 기반 필터링은 사용자와 아이템 사이의 관계를 분석하는 것뿐만 아니라 아이템의 콘텐츠를 분석하여 추

천하기 때문에 다수의 사용자 정보가 필요하지 않아 콜드 스타트 문제에 효율적으로 대응한다. 또한, 콘텐츠 특징을 분석하면서 사용자 개인의 독특한 취향을 고려한 추천이 가능하며 Fig. 2에서 볼 수 있듯 추천의 이유를 설명할 수 있다는 장점이 있다. 하지만 콘텐츠 기반 필터링은 아이템을 분석하여 콘텐츠 특징을 생성하기 어렵다는 단점이 있다. 예를 들어 영화의 경우 장르, 감독, 배우 등의 콘텐츠 특징들을 사람들이 직접 입력해야 하므로 많은 자원이 소모된다는 단점이 있다. 영화가 아닌 온라인 동영상 플랫폼에 사용자들이 직접 올린 동영상에는 사용자들이 콘텐츠 특징을 직접 입력하는 경우가 적고, 콘텐츠 특징들의 카테고리 고리를 정의하기 어려우므로 온라인 동영상 추천을 위한 콘텐츠 특징 정의 및 추출하는 연구가 필요하다.

2. Contents Features used for Video Recommendation System

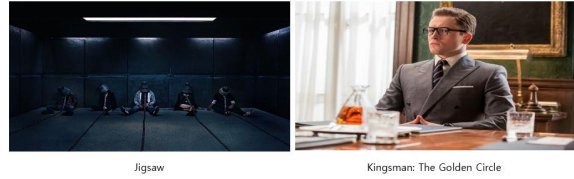
2.1 Visual Features of the Videos

전체 길이의 동영상 또는 영화의 예고편과 같은 하이라이트 편집본에서 자동으로 시각적 특징을 추출하여 사용하는 콘텐츠 기반 추천 시스템에 관한 연구가 진행되고 있다[5, 6]. 영화 추천 시스템에 주로 사용된 동영상의 시각적 특징으로는 조명, 색상, 움직임 및 딥러닝 모델에서 추출한 특정 임베딩 벡터 등이 있다.

영화의 예술적 관점에서 조명은 시청자들이 물체 또는 장소를 보고 어떻게 인식되어야 하는지를 결정하는 요소이다. 밝은 빛으로 그림자가 거의 없는 조명은 코미디나 액션 영화 등에 주로 사용되며 밝은 느낌을 준다. 반면 그늘이 진 듯한 어두운 조명은 섬뜩한 분위기를 조성하며 공포나 스릴러 영화 등에 주로 사용된다. 색감은 장면의 에너지를 나타내며 전반적인 분위기를 결정하는 요소이다. 동영상 속 배우 또는 풍경의 움직임은 동영상의 리듬감을 결정한다. Fig. 3은 영화 속 조명과 색상에 따른 차이를 나타낸다.

딥러닝의 관점에서는 위와 같은 일반적인 동영상의 특징이 아닌 딥러닝 모델을 통해 추출된 임베딩 벡터를 특징으로 한다. 임베딩 벡터는 동영상의 다양한 특징을 포함하고 있다고 판단한다. 이런 시각적 특징들은 영화 및 드라마와 같은 동영상 추천 시스템에서는 좋은 특징으로 작용할 수 있지만, 본 논문의 주제인 교육용 동영상 추천 시스템에서는 좋은 특징이 되기 어렵다.

Movie Lighting Difference



Movie Color Difference

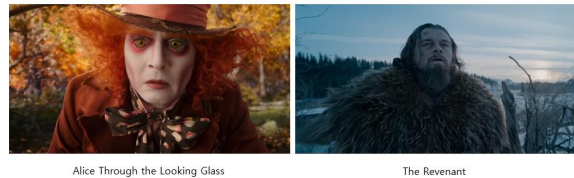


Fig. 3. Example of Visual Features

2.2 Audio Features of the Videos

오디오 특징을 활용하는 추천 시스템 연구는 시각적 특징에서의 조명, 색감 등과 같이 표현 가능한 특징이 아닌 임베딩 벡터의 형태로 동영상의 음성에서 오디오의 특징을 벡터로 변환하여 활용하는 연구가 주를 이룬다[7, 8]. 대표적으로 사용되는 음성 특징으로 블록 레벨 특징(Block-level features; BLF)과 I-벡터 특징(I-vector features), MFCC(Mel-Frequency Cepstral Coefficient) 등이 있다.

블록 레벨 특징은 일반적으로 몇 초간 지속되는 시간의 오디오 세그먼트에서 추출하여 오디오의 시간적 측면을 파악할 수 있는 특징이다. 스펙트럼 측면에서 스펙트럼 패턴, 델타 스펙트럼 패턴, 분산 델타 스펙트럼 패턴의 3가지 패턴과 고조파 측면에서의 상관 패턴, 리듬 측면에서 로그 변동 패턴, 음조 측면에서 스펙트럼 대비 패턴의 총 6가지 패턴의 특징에서 추출되는 고차원 특징 벡터이다.

I-벡터 특징은 음성 처리, 음악 추천, 음향 분석 등 다양한 오디오 관련 분야에서 사용되는 저차원 특징 벡터로 오디오 신호에서 추출된 ID(Identity)의 콤팩트한 통계적 표현이다.

MFCC는 오디오 분야에서 가장 많이 쓰이는 오디오 특징으로 저주파수 대역에서 민감하고, 고주파수 대역에서 상대적으로 둔감한 인간의 청각 시스템과 유사한 특징을 표현한다.

3. Features of Educational Videos

온라인 교육용 동영상의 특징은 화면의 이미지, 화자의 음성, 화자의 제스처 등의 기호를 포함하며 텍스트의 정보뿐만 아니라 비언어적인 정보도 함께 제공한다. 교육용 동영상의 비언어적인 정보는 학습자에게 좀 더 명확한 이해가 가능하도록 돕는 역할을 한다.

커뮤니케이션 이론 중 메라비언의 법칙(The Law of Mehrabian)은 한 사람이 상대방으로부터 받는 이미지는 언어가 7%, 비언어적 요소가 93% 영향을 받아 결정된다는 법칙이다. 메라비언은 효과적인 의사소통에 있어 말보다 비언어적인 요소가 더 중요하다고 한다[9]. 메라비언의 법칙은 비언어적인 요소는 시각 55%와 청각 38%로 나누어진다. 언어는 말의 내용을 의미하고, 시각은 교수자의 자세, 용모, 복장, 제스처 등 외적으로 보이는 부분을 말하며, 청각은 교수자의 목소리 톤이나 음색처럼 언어의 품질을 의미한다.

온라인 교육용 동영상은 일반적인 대면 강의와 비교하여 교수자의 표정이나 제스처, 자세 등 비언어적인 요소를 표현하는데 제한적이기 때문에 다양한 멀티미디어 요소를 활용한다. 온라인 교육용 동영상의 멀티미디어 요소는 크게 시각적 요소와 청각적 요소로 구분되며 시각적 요소에는 자막, 사진 등이 포함되고, 청각적 요소에는 소리, 효과음 등이 포함된다.

온라인 강의의 콘텐츠 특징에 따른 영향력에 관한 연구는 콘텐츠 특징을 시각적 요소와 청각적 요소로 나누어 접근하고, 각 요소가 학습자의 학습성취에 영향을 미치는 정도를 조사한다[10-12]. 학습 참여에 영향력이 높은 요소로 시각적 요소에는 교수자가 화면을 차지하는 비율과 자막의 유무, 청각적 요소에는 교수자의 목소리의 특징으로 말의 속도, 높낮이, 크기, 느낌, 마이크 사용 여부가 있다. 본 논문에서는 교수자를 기준으로 멀티미디어 요소를 콘텐츠 특징으로 추출하여 온라인 교육용 동영상 추천 시스템에 적용한다. 시각적 특징으로 기존 연구에서 사용한 특징인 화면 속 교수자의 등장 여부와 화면 속 교수자 크기의 비율, 자막 여부를 사용한다. 학교 교육 현장에서 교사가 가장 많이 사용하는 매체는 교사의 목소리이다. 교사의 목소리에 관한 연구는 학생들이 시각적 단서 없이 목소리만으로 화자의 성, 나이, 성격, 건강, 수업 방식 등을 유추할 수 있으며, 교사의 목소리가 내용전달 및 학생의 집중도에 구체적인 영향을 미친다. 또 최근 교육 현장에서는 교수자의 방언 사용 여부도 문제가 되는 등 본 논문에서 교육용 동영상 추천 시스템에 활용할 청각적 특징으로 교수자의 음성에서 성별, 나이, 방언 사용 여부를 사용한다.

III. The Proposed Scheme

1. Educational Video Content Features

교육용 동영상 추천의 관점에서 사용자에게 중요한 동영상의 콘텐츠 특징을 추출한다. 사용자가 교육용 동영상을 시

청하기 위해 검색어를 입력했을 때, 사용자가 원하는 교육 내용(검색어)과 실제 동영상의 내용이 얼마나 일치하는지가 중요하다. 또한, 교육용 동영상에서 화자는 사용자의 선택에 있어 중요한 요소이다. 화자의 음성은 화자에 대한 사회, 언어학적 정보를 포함하고 있어 사용자에게 선호되는 정보를 구분할 수 있다. 본 논문에서는 화자의 음성 특징 중 화자의 성별, 나이, 방언 사용 여부를 추출하여 이를 사용자의 선호 특징 정보를 활용하는 추천 시스템을 제안한다.

1.1 Semantic Similarity

동영상의 제목은 동영상의 내용을 요약하고 대변하는 하나의 특징으로 검색어와 동영상의 내용이 얼마나 유사한 의미를 지니고 있는지 측정하여 이를 추천 시스템에 사용하는 특징 중 하나로 사용한다.

두 문장 사이의 의미론적 유사도는 두 문장이 서로 다른 단어와 문장 순서를 가짐에도 얼마나 비슷한 정보를 찾고 있는지 또는 비슷한 대답이 예상되는지 등을 판단하는 지표로 단어 그대로 두 문장이 얼마나 의미론적으로 유사한지 판단한다[13]. 일반적으로 텍스트 임베딩을 사용하여 의미론적 유사도를 측정하는데 문장을 임베딩 모델을 통해 벡터화하고 벡터화된 문장 사이의 벡터 유사도를 코사인 유사도를 사용하여 계산한다.

본 논문에서는 문장을 벡터화하기 위한 텍스트 임베딩 모델 중 BERT(Bidirectional Encoder Representations from Transformers)를 사용하여 검색어와 동영상의 제목 사이의 의미론적 유사도를 측정한다. BERT는 구글에서 개발한 NLP(자연어처리) 사전 훈련 기술로 모든 자연어처리 분야에서 좋은 성능을 내는 언어 모델이다[14].

1.2 Speaker Appearance

교육용 동영상에서 화자의 등장 여부는 사용자가 동영상에 집중하는 데 있어 중요한 특징이 될 수 있다. 교육용 동영상 화면에서 화자가 직접 등장하여 손짓, 몸짓을 섞어가며 이야기할 때 좀 더 집중이 잘 되는 경우가 있고, 화면에서 화자가 등장하지 않고 음성만으로 전달하는 것이 더 집중이 잘 될 수 있다.

화자의 출연 여부를 확인하기 위해 객체 검출 딥러닝 모델로 one-stage 방법으로 설계되어 매우 빠른 속도로 객체를 검출한다는 장점이 있는 YOLO(You Only Look Once)를 사용한다[15]. one-stage 방법은 객체를 인식할 때 객체의 위치를 찾고, 그 객체가 무엇인지 예측하는 과정을 동시에 해결하는 방법이다.

1.3 Speaker Gender, Age, Dialect

화자의 음성 특징으로 화자의 성별, 나이, 방언 정보를 추출한다[16-18]. 오디오 신호 처리 분야 및 화자 인식 분야에서 널리 사용되는 오디오 신호에서 추출한 특징값인 MFCC를 사용하여 위 특징을 추출하는 딥러닝 모델을 설계한다. 딥러닝 모델의 구조는 다음 Fig. 4와 같다. 교육용 동영상의 시작 부분은 BGM 또는 공백으로 화자의 목소리를 검출하기 어려워 동영상의 재생시간 10초 이후의 구간에서 임의로 4초의 오디오를 추출하여 13차 MFCC 오디오 특성을 추출하여 4초간의 평균값을 입력으로 사용한다. 모델은 5개의 은닉층과 분류층으로 구성하며 은닉층별 크기는 각각 256, 256, 128, 128, 64로 구성하고 각 은닉층은 0.3 확률로 DropOut 과정을 거쳐 오버피팅을 방지한다. 마지막으로 분류층은 교육용 비디오 추천에 맞게 세분화된 분류 대신 성별의 경우 남성과 여성, 나이의 경우 일반 성인과 그 외 나이대, 방언의 경우 표준어 또는 방언을 사용 여부에 대해 이진 분류를 한다.

본 논문에서는 AI Hub에서 제공하는 한국인 대화 음성 데이터 세트를 사용하여 화자 특징으로 성별, 나이, 방언 정보를 추출한다[19]. 한국인 대화 음성 데이터 세트는 나이, 원거리, 노이즈, 다자대화 등 다양한 환경을 고려하여 한국인의 일상 대화를 인식하고 음성을 문자로 변환하기 위한 데이터 세트로 약 4,000시간의 음성 데이터와 약 400만 문장의 텍스트 데이터로 이루어져 각 음성 데이터에 대해 화자의 성별, 나이, 거주 지역, 방언 정보 및 음성 대화의 주제와 음질 정보 등의 메타데이터 정보를 제공한다. 본 논문에서는 실험을 위해 한국인 대화 음성 데이터 세트에서 음성 데이터 중 음질이 정상인 데이터에서 성별, 세대, 화자 방언 라벨링 정보를 활용하여 화자 음성 특징 추출 모델을 훈련한다.

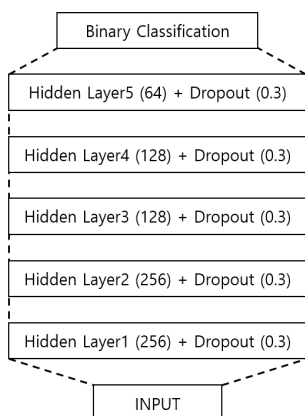


Fig. 4. Audio Feature Extraction Model

Table 1. Extraction Accuracy by Speaker's Audio Features

Audio Features	Accuracy
Speaker Gender	92.20 %
Speaker Age	86.69 %
Speaker Dialect	83.45 %

한국인 대화 음성 데이터 세트에서 화자 성별 추출을 위해 남성과 여성의 음성 데이터를 각각 10,000개씩 임의로 추출하고, 화자 나이 추출을 위해 일반 성인의 음성 데이터를 약 13,000개, 그 외 유아, 청소년, 고령층의 데이터를 약 9,000개를 임의로 추출하고, 화자 방언 특징 추출을 위해 경기/서울 지역의 음성 데이터를 약 9000개, 제주를 제외한 강원, 충청, 전라, 경상 지역의 방언 음성 데이터를 약 8000개를 임의로 추출한다. 화자의 음성 특징에 따라 서로 다르게 추출된 데이터들을 7:3의 비율로 훈련, 테스트 데이터 세트를 분리하여 앞서 설계한 딥러닝 모델을 특징 추출을 위한 분류기로써 훈련하고 그 성능을 측정한다. 동영상 속 화자의 음성 특징을 이진 분류를 통해 추출하기 때문에 정확도만 사용하여 성능을 측정한다. 다음 Table. 1은 화자의 음성 특징별 추출 정확도를 나타낸다.

화자 성별 분류에서 92.20%의 정확도를 보이고, 화자 나이 분류에서 86.69%의 정확도를 보이고, 화자 방언에서 83.45%의 정확도를 보인다. 화자 나이 및 방언을 일반적인 상황과 그렇지 않은 상황의 이진 분류를 통해 특징 추출의 정확도를 높여 사용자의 선호 정보를 파악한다.

1.4 Youtube Data

사용자가 원하는 최적의 교육용 동영상을 추천하기 위해 유튜브에서 제공하는 정보를 바탕으로 동영상의 자막 유무, 화자의 말하기 속도, 동영상의 길이 및 게시일을 특징으로 추출한다.

교육용 동영상에서 화자가 방언을 사용하거나 동영상 촬영시 환경적인 요인 등으로 사용자가 화자의 말을 잘 이해하지 못할 수 있다. 이를 보완해줄 수 있는 것이 업로더가 직접 올린 자막의 유무이다. 자막이 있을 때 교육적인 측면에서 사용자는 더욱 효과를 볼 것으로 기대한다. 유튜브의 경우 자체적으로 자막을 생성해주긴 하지만 한국어의 경우 부족한 성능을 보여 업로더가 직접 업로드한 자막을 기준으로 한다.

화자의 말하기 속도 또한 사용자에게 있어 중요한 요소지만 화자의 말하기 속도를 측정하기 위해 전체 동영상을 분석하는 일은 많은 자원이 소모되는 어려운 문제이기 때문에 본 논문에서는 자막에서 생성된 텍스트의 길이를 동영상의 길이로 나누어 분당 글자 수를 기준으로 말의 빠르기를 특징

화한다. 분당 글자 수가 400자 이상일 때 빠름, 350~400자 일 때 보통, 350자 이하일 때 느림으로 표현한다.

동영상에서 게시일과 길이 또한 추천 시스템에 적용할 특징으로 충분하다. 사용자가 원하는 내용 이외의 내용까지 포함하는 교육용 동영상의 경우 사용자가 예측하는 시간보다 길어질 수 있고 너무 짧은 동영상은 원하는 내용을 충분히 포함하지 못했다고 판단할 수 있다. 유튜브를 통해 사용자가 교육용 동영상을 검색할 때 인터넷 강의와 같은 유형보다 특정 교육 주제에 대한 검색이 많을 것이라는 가정하에 10분 이하, 10분~20분 사이, 20분 이상으로 분류한다. 또 일반적으로 교육용 동영상의 내용은 불변하는 경우가 많지만 최근 트렌드를 반영되어 내용이 변하는 교육 분야도 있으므로 다양한 기준인 3개월, 6개월, 1년, 3년으로 분류한다.

2. Method of Online Educational Video Recommendation

2.1 Rank Reordering using User Preference Features

본 논문에서는 콜드 스타트 상황에서 사용자의 선호 특징을 조사하고, 유튜브에서 파악할 수 있는 메타데이터 정보를 활용해 검색 결과를 순위 재정렬하는 지표로 사용될 점수를 다음 식 (1)과 같이 정의하고 순위 재정렬 점수라고 정의한다. 의미론적 유사도는 BERT 모델을 이용하여 검색어와 검색 목록의 동영상 제목 사이의 유사도를 측정 한 값으로 한다. 유튜브 인기도는 유튜브에서 제공하는 메타데이터를 이용하여 조회 수 대비 얼마만큼의 유저 피드백이 있었는지를 판단하기 위해 본 논문에서 설정한 지표로 식 (2)의 방법으로 계산한다. 또한, 선호 점수는 Fig. 5의 특징들에 대해 사용자가 선택한 값이 '상관없음' 일 때 0점을 부여하고, 사용자가 선택한 값과 동영상의 값이 서로 같을 때 1점을 주고, 다르면 -1점을 부여하여 그 합계를 계산하여 [-1, 1] 사이로 정규화한 값이다.

$$\text{순위 재정렬 점수} = \text{의미론적 유사도} + 10 \times \text{유튜브 인기도} + \text{선호 점수} \quad (1)$$

$$\text{유튜브 인기도} = \frac{\text{좋아요} + \text{싫어요} + \text{댓글 수}}{\text{조회수}} \quad (2)$$



Fig. 5. Educational Video Content Features

2.2 Rank Reordering using User Viewing History

교육용 동영상의 여러 특징을 추출하여 이를 와이드 앤 딥 추천 시스템에 적용하여 사용자에게 개인화된 추천을 제공한다. 추천 시스템은 콜드 스타트에서 사용자의 검색 결과와 피드백 결과를 기록하고, 사용자의 긍정적인 피드백을 가진 데이터들을 기록한다. 데이터가 쌓였을 때 시스템은 와이드 앤 딥 알고리즘을 이용해 추천을 제공한다. Fig. 6은 본 논문에서 사용한 와이드 앤 딥 모델의 구성을 의미한다. 와이드 앤 딥(Wide & Deep) 추천 시스템은 구글 플레이에 적용된 알고리즘으로 추천의 대상이 되는 아이템에 대한 정보로 사용자의 검색 결과에 대해 순위를 재정렬한다. 와이드 앤 딥 모델은 와이드 모델의 상호작용 기억(Memorization of Interaction) 개념과 딥 모델의 일반화(Generalization) 개념을 합쳐 범용적으로 추천을 해 주는 딥러닝 기반 추천 알고리즘 모델이다[20].

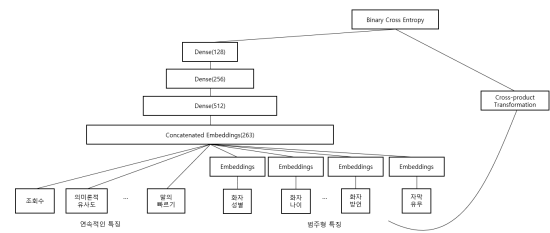


Fig. 6. Wide & Deep Recommendation Model

동영상의 특징에 따라 연속성을 가지는 특징과 범주형의 특징을 분리하여 사용한다. 유튜브 동영상에서 추출한 조회 수, 좋아요 수, 싫어요 수, 댓글 수, 1분당 글자 수, 의미론적 유사도, 게시일 및 영상 내 화자의 등장 비율, 말의 빠르기 특징들은 연속적인 특징으로 특징 자체의 값을 그대로 입력으로 사용한다. 그 외 범주형 특징인 화자의 성별, 나이, 방언 및 지막 유무의 특징들은 숫자로 인코딩한다. 숫자로 표시된 특징을 정규화하고 그 값을 다시 Polynomial Feature로 변경하여 다항 회귀(Polynomial Regression)의 방법으로 검색 기록 데이터를 입력으로 사용자가 좋다고 표시할 동영상을 예측하도록 모델을 훈련한다. 이때, 다항 회귀는 비선형 학습 데이터를 다항식 함수로 변형하여 선형적인 특징을 가지게 함으로써 선형 회귀를 확장하는 방법이다.

모델의 입력으로 유튜브 검색 결과 상위 10개와 각 동영상의 특징을 사용한다. 연속적인 특징과 범주형 특징 모두를 딥 모델의 입력으로 사용하고, 추가로 범주형 특징은 와이드 모델의 입력으로 사용하여 벡터 곱 변환 연산을 통해 특징을

추출한다. 딥 모델과 와이드 모델을 동시에 훈련하여 유튜브 검색 결과를 사용자의 취향에 맞게 재정렬한다.

IV. Recommendation System Implementation

1. Design of Online Educational Video Recommendation System

본 장에서는 온라인 플랫폼 중 유튜브를 기반으로 교육용 동영상 추천 시스템을 설계하고 웹 애플리케이션으로 구현한다. 동영상에서 추출한 특징으로 사용자의 선호도를 조사하여 콜드 스타트 문제에 대응하는 순위 재정렬 방법을 적용하고, 사용자의 기록이 쌓이면 이들 특징을 기반으로 딥러닝 추천 알고리즘에 적용한다.

Fig. 7은 본 논문이 제안하는 온라인 플랫폼 중 유튜브에서 교육용 동영상 추천 시스템의 개념도를 나타낸다. 먼저 사용자가 교육용 동영상을 찾기 위해 검색어를 입력하면, 추천 시스템은 유튜브에서 검색어에 대한 동영상 목록을 가져온다. 사용자는 선호 정보 이용 추천 방법이나 시청 기록 기반 추천 방법을 사용하여 동영상 정렬 목록을 받는다. 사용자의 선호 정보로 본 논문에서 제안하는 선호 점수를 측정하고 이를 기준으로 유튜브 검색 결과를 내림차순으로 순위를 재정렬하여 사용자에게 제공한다. 사용자가 시청 기록 기반 추천 방법을 선택한 경우 사용자의 시청 기록의 동영상에서 추출된 동영상의 콘텐츠 특징과 유튜브가 제공하는 동영상의 특징으로 와이드 앤 딥 추천 알고리즘을 훈련하여 사용자 맞춤형 동영상을 추천한다. 최종적으로 시스템이 추천한 동영상 중 사용자가 만족한 동영상을 체크하여 사용자 기록에 추가한다.

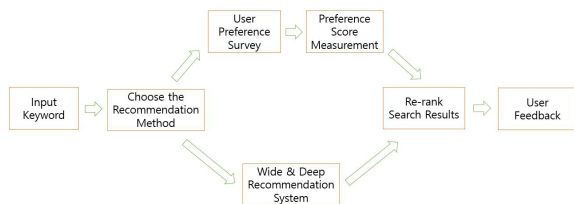


Fig. 7. Conceptual Diagram of Educational Video Recommendation System

2. Implementation of Online Educational Video Recommendation System

유튜브 기반 교육용 동영상 추천 시스템을 웹 애플리케이션으로 구현하기 위해 Python 언어와 Tensorflow,

Streamlit 라이브러리를 사용한다. Tensorflow는 구글에서 만든 라이브러리로 딥러닝을 쉽게 구현할 수 있도록 기능을 제공한다. Streamlit은 2019년 등장한 Python 기반의 웹 애플리케이션 툴로 데이터 사이언스/머신러닝 및 딥러닝을 배포하는데 유용한 기능을 제공한다[21].

Fig. 8은 추천 시스템을 구현하여 사용하여 기존 유튜브와 비교한 예시이다. 검색어로 ‘파이썬 조건문’을 입력하고 왼쪽 상단에서 두 가지 추천 방법 중 한 가지를 선택한다. 선호 정보 조사 기반 추천 방법은 선호 정보 조사를 통해 사용자가 원하는 특징을 선택한다. 시스템은 이를 바탕으로 기존 유튜브에서의 검색 결과에서 선호 정보 점수를 기준으로 검색 결과의 순위를 재정렬하여 사용자에게 추천한다. 사용자는 제공된 동영상을 시청하고 좋았을 경우 동영상 하단에 체크 박스를 통해 좋았다는 표시를 하여 사용자 선호 동영상 시청 기록에 저장한다.

두 번째 추천 방법인 시청 기록 기반 추천 방법은 체크 박스를 통해 유저의 시청 기록 데이터가 쌓이면 동영상의 콘텐츠 특징과 유튜브가 제공하는 특징으로 와이드 앤 딥 추천 알고리즘을 훈련하여 사용자의 선호도 조사 없이 사용자의 취향을 판단해 개인화된 유튜브 검색 결과를 순위 재정렬하여 사용자에게 제공한다.

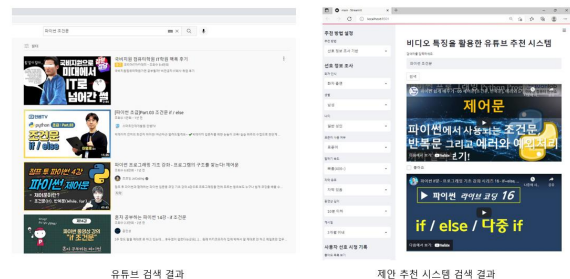


Fig. 8. Result of Implementation of the Online Education Video Recommendation System

V. Conclusions

본 논문에서 웹 애플리케이션으로 유튜브 기반 교육용 동영상 추천 시스템을 구현하고 사용자 만족도 조사를 통해 평가한다. 구현한 유튜브 기반 교육용 동영상 추천 시스템의 사용자 만족도를 조사하기 위해 4가지 항목에 대해 설문 조사를 진행하였다. 설문 조사는 3일 이상 추천 시스템을 사용한 20~30대 남녀 사용자 41명을 대상으로 Fig. 9와 같이 구글 설문지를 통해 진행하였다.

사용자 41명의 설문 조사 중 전반적인 시스템에 대한 평가 항목인 2, 5, 6, 7번째 항목을 각각 ‘선호 정보 조사

의 적절성', '사용자 기록 기반 개인화 추천 만족도', '추천 시스템 추천 성능 만족도', '추천 시스템 편의성 만족도'로 표기한 설문 조사 결과는 다음 Fig. 10과 같다.

만족도를 전체 응답 수 대비 보통 이상 응답 수로 정의하고 각 항목에서의 만족도 평균을 전체 평균 만족도로 정의하여 추천 시스템을 평가한다.

'선호 정보 조사의 적절성' 항목에서 87.8%의 만족도를 기록했다. 대체로 선호 정보 조사 항목이 적절하다는 사용자들의 평가가 있었지만, 추가로 유튜브가 제공하는 동영상 특징인 조회 수, 동영상 채널의 구독자 수 등의 항목이 있었으면 좋겠다는 응답이 있었다.

'사용자 기록 기반 개인화 추천 만족도'에서 만족도는 78.04%로 다른 항목에 비해 사용자들의 만족도가 낮다. 사용자 기록 기반 추천은 사용자의 기록이 많을수록 좋은 추천을 제공하는데 사용자들의 기록이 충분히 쌓일 만큼 사용하지 않고 설문 조사를 진행하여 다소 아쉬운 만족도 결과가 나온 것으로 보인다.

'추천 시스템 추천 성능 만족도'에서 85.36%의 만족도로 두 가지 방법의 추천 방법 모두에서 사용자를 만족시키고 있다고 판단할 수 있다.

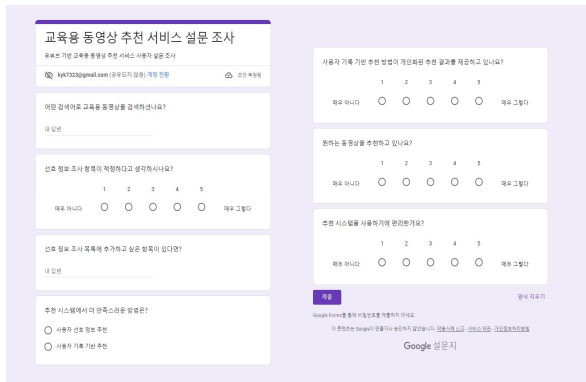


Fig. 9. Recommendation System Satisfaction Survey

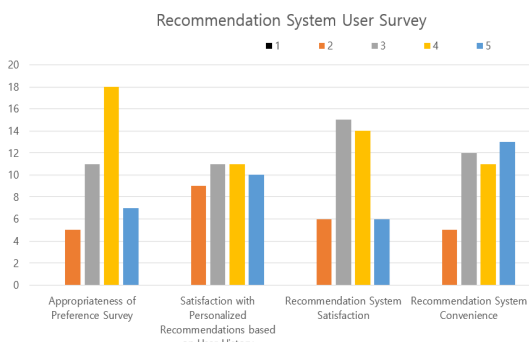


Fig. 10. Recommendation System User Survey

'추천 시스템 편의성 만족도'에서 87.80%로 구현한 웹 애플리케이션이 사용자 편의성에 있어 좋은 UI를 구축한 것으로 판단할 수 있다.

본 논문이 제안하는 추천 시스템은 사용자의 두 가지 선택에 따라 다른 추천 방법으로 교육용 동영상은 추천한다. 온라인 교육용 동영상의 콘텐츠 특징을 정의하고, 추출하여 추천 시스템에 활용한다. 사용자의 선호 정보만 사용하여 추천하는 방법과 사용자의 시청 기록을 기반으로 추천하는 방법 모두 사용자들로부터 좋은 평가를 받았지만, 동영상의 콘텐츠 특징을 확대하여 유튜브가 제공하는 기본적인 동영상의 특징도 포함되었으면 좋겠다는 의견에 따라 선호 정보를 확대하는 연구가 필요하다. 또한, 동영상의 콘텐츠 특징으로 화자의 음성 특징뿐만 아니라 영상 등에서도 추출하여 추천 시스템에 활용하는 연구가 필요하다.

REFERENCES

- [1] Banronbodo.com, '1 minute on the Internet... 500 hours of YouTube uploaded', <http://www.banronbodo.com/news/articleView.html?idxno=10153>
- [2] Nasmedia, 'Internet user survey NPR 2021', <https://www.nasmedia.co.kr/NPR/2021/>
- [3] Javed, U., Shaukat, K., Hameed, I. A., Iqbal, F., Alam, T. M., & Luo, S. A review of content-based and context-based recommendation systems. *International Journal of Emerging Technologies in Learning (iJET)*, 2021, 16(3), 274-306.
- [4] Reddy, S. R. S., Nalluri, S., Kuniseti, S., Ashok, S., & Venkatesh, B. Content-based movie recommendation system using genre correlation. In *Smart Intelligent Computing and Applications*, 2019. Springer, Singapore, pp. 391-397.
- [5] Deldjoo, Y., Schedl, M., Cremonesi, P., & Pasi, G. Recommender systems leveraging multimedia content. *ACM Computing Surveys (CSUR)*, 2020, 53(5), 1-38.
- [6] Deldjoo, Y., Di Noia, T., Malitesta, D., & Merra, F. A. Leveraging Content-Style Item Representation for Visual Recommendation. In *European Conference on Information Retrieval*, 2022, Springer, Cham, pp. 84-92.
- [7] Deldjoo, Y., Constantin, M. G., Eghbal-Zadeh, H., Ionescu, B., Schedl, M., & Cremonesi, P. Audio-visual encoding of multimedia content for enhancing movie recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems*, 2018, September, pp. 455-459.
- [8] Deldjoo, Y., Schedl, M., Hidasi, B., & Knees, P. Multimedia recommender systems. In *Proceedings of the 12th ACM Conference on Recommender Systems* 2018, September, pp.

537-538.

- [9] Mehrabian, A. (1971). Silent messages (Vol. 8, No. 152, p. 30). Belmont, CA: Wadsworth.
- [10] Kang, Jae Shin, Kang, Tae Yim. Analysis of Content Components that Influence Learning Participation in Online Lectures for Multimedia Practice Subjects. Journal of the Korean Society of Design Culture, 2020, 26(4), 19-32.
- [11] Kim, Jong Moo. Analysis of learner's preference on the type of e-learning lecture contents -About basic linguistics subject and theoretic subject-, Korea Design Knowledge Society, 2015, 34, 175-182.
- [12] Namin Shin, Jung Hoon Lee. Students' emotional responses to teachers' voice media and preferred voice features of the teachers, Korean Society Of Educational Technology, 2009, 25(4), 29-52.
- [13] Chandrasekaran, D., & Mago, V. Evolution of semantic similarity –a survey. ACM Computing Surveys (CSUR), 2021, 54(2), 1-37.
- [14] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. 2018. arXiv preprint arXiv:1810.04805.
- [15] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779-788.
- [16] Sweta, S., Babu, J. M., Palempati, A., & Kanhe, A. Cepstral Coefficient-Based Gender Classification Using Audio Signals. In Sustainable Advanced Computing, 2022. Springer, Singapore. pp. 81-90.
- [17] Grzybowska, J., & Kacprzak, S. Speaker Age Classification and Regression Using i-Vectors. In INTERSPEECH. 2016, September, pp. 1402-1406.
- [18] Lulu, L., & Elnagar, A. Automatic Arabic dialect classification using deep learning models. Procedia computer science, 2018, 142, 262-269.
- [19] AI Hub, Korean Conversation Voice, <https://aihub.or.kr/aidata/7968>.
- [20] Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., ... & Shah, H. Wide & deep learning for recommender systems. In Proceedings of the 1st workshop on deep learning for recommender systems, 2016, September, pp. 7-10.
- [21] Streamlit library, <https://streamlit.io/>

Authors



Young Kook Kim received the B.S. degree in Mathematics from Soongsil University (SSU), Korea, in 2017. He is currently a M.S. Student in the Department of Software, Soongsil University (SSU).

His research interests are Deep Learning, Machine Learning, Recommendation System, NLP and Big data analysis.



Myung Ho Kim received the B.S. in Department of Computer Science and Engineering from Soongsil University, Korea, in 1989. M.S. and Ph.D. degrees in Department of Computer Engineering from

Postech University, Korea, in 1991 and 1995, respectively. He is currently a professor in the Dept. of Software, Soongsil University. He is interested in Machine Learning, Deep Learning and Block chain.