교육훈련생의 숙련 등급을 고려하는 콘텐츠 추천 시스템

황상호*, 손창식*, 진상현*, 강원석*
*대구경북과학기술원 지능형로봇연구부
e-mail: wskang@dgist.ac.kr

Content recommendation system considering skill level of trainees

Sang-Ho Hwang*, Chang-Sik Son*, Sang-Hyun Jin*, Won-Seok Kang*
*Division of Intelligent Robot, DGIST

요 약

본 논문에서는 교육 훈련생의 교육 상태에 맞는 콘텐츠를 추천해주는 교육 콘텐츠 추천시스템을 제안한다. 제안하는 시스템은 데이터베이스에 저장되어 있는 콘텐츠의 텍스트를 활용하여 콘텐츠를 등급별로 분류하는 콘텐츠 분류 기능부분과 교육 훈련생의 등급 및 교관 등에 의해 설정되는 필터링 정보를 활용하여 교육 훈련생에 맞는 콘텐츠 추천리스트를 생성하는 콘텐츠 추천 기능으로 구성되어있다. 실험에서는 콘텐츠 분류기능에 대한 정확도 평가를 수행하였고, ANN 평가모델로 평균 65%의 정확도를 보였다.

1. 서론

교육 과정에 있어 교육 훈련생에 교육 상태에 따른 다음 교육 콘텐츠를 추천해주는 것은 교육 훈련생의 흥미, 교육의 효율성을 위해 중요하다[1-2]. 교육, 영화, 드라마, 음악, 뉴스 등의 콘텐츠에 대한 추천은 대체적으로 사용자의 성향 또는 과거 사용이력 등의 정보를 활용하여 이루어진다[3]. 이러한 추천 시스템에 사용되는 대표적인 추천 기술로는 협업 필터링 기법이며, 콘텐츠를 사용하는 사용자들의 행동 패턴을 분석하여 추천한다. 또한 이 기법은 콘텐츠 추천을 원하는 사용자와 유사한 행동 성향을 보였던 과거 사용자들의 추천 사례들을 이용하여 서비스를 제공하는 것이 주된 특징이라 할 수 있다. 그러나 새로운 콘텐츠가 추가될 경우 추천에 사용될 사례들이 없어 해당 콘텐츠의 서비스가 불가능하다는 제약점이 있다[4-6].

본 논문에서는 교육 훈련생의 교육 수준에 적합한 콘텐츠를 제공하기 위해, 콘텐츠 기반 필터링 기법을 기반으로 한 교육 콘텐츠 추천시스템을 제안한다. 제안된 시스템은 데이터베이스에 저장되어 있는 콘텐츠의 텍스트를 활용하여 콘텐츠를 적절한 등급으로 분류하는 콘텐츠 분류 기능부분과 교육 훈련생의 등급 및 교관 등에 의해 설정되는 필터링 정보, 그리고 이들 2가지 기능을 활용하여 교육 훈련생에 맞는 콘텐츠 추천리스트를 생성하는 콘텐츠 추천 기능으로 구성되어있다.

2. 교육 콘텐츠 추천 시스템

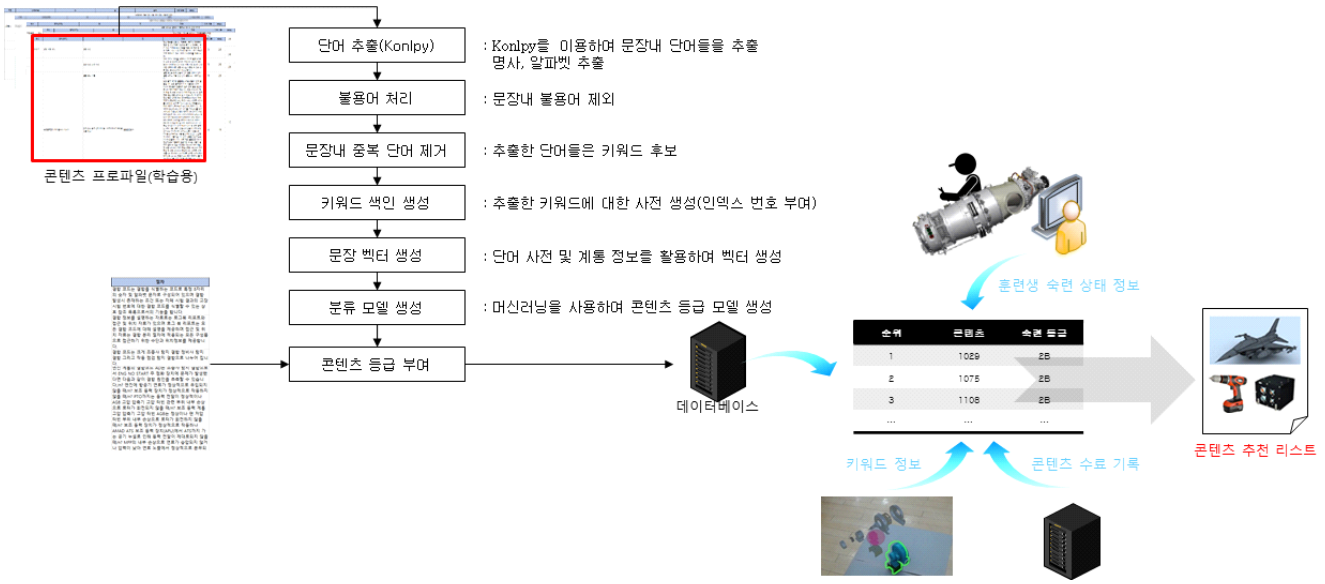
제안된 교육 콘텐츠 추천 시스템은 콘텐츠를 등급별로 분류하는 기능과 교육 훈련생의 교육 수준에 맞는 교육 콘텐츠를 추천하는 기능으로 구성되어 있다.

1. 콘텐츠 분류 기능

콘텐츠 분류 기능은 콘텐츠를 텍스트(제목, 교육 설명, 교육 절차 등)에 기반을 두어 적절한 등급으로 분류한다. 이 기능은 텍스트 불러오기, 단어 추출, 벡터화를 거쳐 생성된 문장특징과 머신러닝을 활용하여 만들어진 분류모델을 기반을 두어 이루어진다.

콘텐츠를 적절한 등급으로 분류하기 위해, 제안하는 시스템은 우선 데이터베이스에 저장되어있는 콘텐츠의 교육과 관련된 텍스트들을 읽어온다. 불러오는 콘텐츠 텍스트는 제목, 교육 내용, 교육 절차 등이 포함되어 있다. 각 콘텐츠의 텍스트들은 한국어 정보처리를 위한 패키지인 KoNLPy[7]의 KKMA, OKT, KOMORAN, HANNANUM과 같은 형태소 분석기를 활용하여 모든 텍스트를 단어 단위로 추출한다. 추출된 단어들은 품사가 태깅되어 있고, 본 논문에서는 추출된 단어들 중에 의미전달이 강한 일반명사, 고유명사 및 외국어만 활용하고 나머지 품사에 해당하는 단어들은 텍스트에서 제외한다. 또한 학습 및 벡터 생성에 불필요한 단어들은 불용어로 등록하여 문장에서 제거한다. 키워드에 기반을 둔 문장 벡터 생성을 위해, 단어 단위로 추출된 문장 내에 중복 단어는 제거된다.

추출된 각 단어에는 고유한 인덱스 번호가 부여되며, 단어들의 인덱스 번호를 활용하여 문장에 대한 벡터를 만든다. 문장에 대한 벡터 생성에는 TF-IDF(Term



(그림 1) 교육 콘텐츠 추천 시스템의 구조

Frequency-Inverse Document Frequency)를 활용하였다. 콘텐츠 분류에 활용될 모델은 ANN(Artificial Neural Network), kNN(k-Nearest Neighbor)과 같은 머신러닝 알고리즘을 활용하여 생성된다.

2. 콘텐츠 추천 기능

제안된 콘텐츠 추천시스템에서 추천 기능은 다음의 절차에 따라 수행된다. 우선 데이터베이스로부터 교육 훈련생이 수행하고 있는 교육 커리큘럼과 훈련생의 숙련 등급을 불러온다. 교육 훈련생의 과거 콘텐츠 수강 또는 수료 이력과 교관 또는 교육 훈련생에 의해 입력되는 교과 키워드를 활용하여 교육 콘텐츠 추천 리스트를 생성한다. 생성한 콘텐츠 추천 리스트는 추후 교육 훈련에 활용할 수 있도록 데이터베이스에 저장한다. 제안하는 콘텐츠 추천 시스템은 이러한 과정을 통해 교육 훈련생의 숙련 등급에 적절한 교육 콘텐츠를 추천할 수 있다. 그림 1은 제안한 교육 콘텐츠 추천 시스템의 구조를 보여준다.

3. 콘텐츠 분류 모델의 정확도 평가

제안하는 콘텐츠 분류 모델의 정확도를 측정하기 위해서 항공정비 교육을 위해 사용되는 교본을 활용하여 정확도를 평가하였다. 실험 평가를 위해 사용된 콘텐츠의 수는 4917개이다. 표 1은 실험에서 활용한 데이터 정보를 보여준다. 교본의 숙련 등급은 수행 난이도에 따라 12등급으로 분류된다.

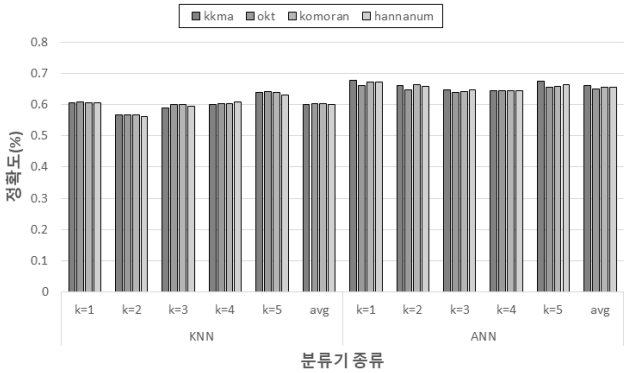
<표 1> 콘텐츠 등급별 개수

등급	1A	1B	1C	2A	2B	2C	3A	3B	3C	4A	4B	4C
개수	239	43	55	160	159	34	32	186	172	570	607	2656

<표 2> 평가모델의 실험 매개변수

분류기 종류	실험 매개변수
kNN	k = 5
ANN	활성화 함수: ReLU[8] 은닉층 수: 3 (각 층의 노드 수 1024) 최적화 기법 : Adam[9]

형태소 분석기는 KoNLPy를 사용하였으며, 일반명사, 고유명사, 외국어 품사에 해당하는 단어만 활용하였다. 모든 형태소 분석기는 별도의 사전을 등록하지 않았다. 객관적인 통계적 성능평가 결과를 제공하기 위해서, 본 실험에서는 계층적 5-겹 교차검증(stratified 5-fold cross validation)을 사용하였다. 또한 실험에서 사용한 평가모델로는 kNN, ANN이며 각 평가모델에서 사용된 실험적 매개변수는 표 2와 같다.



(그림 2) 콘텐츠 분류 모델의 정확도 평가

그림 2는 콘텐츠 분류 모델의 정확도 평가 결과를 보여준다. 실험평가에서는 KoNLPy의 4종의 형태소 분석기(KKMA, OKT, KOMORAN, HANNANUM) 모두 활용하여 정확도 비교를 하였고, kNN 평가모델을 사용한 정확

도 분석에서는 4종의 형태소 분석기에서 큰 차이를 보이지 않았다. ANN알고리즘의 경우 KKMA 66%, OKT 64%, KOMORAN 65%, HANNANUM 65%로 KKMA 형태소 분석기를 사용한 경우 가장 좋은 분류 정확도를 보였다. 사용된 평가모델별 정확도 성능에서는 kNN이 평균 60%, ANN이 66% 수준이었고, ANN 평가모델이 콘텐츠 추천기로 적절함을 볼 수 있었다.

4. 결론

본 논문은 교육 훈련생의 숙련 등급에 맞는 콘텐츠를 추천해주는 콘텐츠 추천 시스템에 대한 설계에 대하여 제안하였다. 제안하는 추천 시스템은 콘텐츠의 텍스트를 활용하여 콘텐츠를 등급별로 분류하는 기능부분과 교육 훈련생의 등급 및 교관 등에 의해 설정되는 필터링 정보를 활용하여 교육 훈련생에 맞는 콘텐츠 리스트를 생성하는 콘텐츠 추천 기능으로 구성되어있다. 제안하는 추천 시스템은 두 가지 기능을 활용하여 교육 훈련생의 교육 훈련 상태에 적절한 교육 콘텐츠를 추천해줄 수 있다. 다만, 구현에 활용한 KoNLPy의 경우 기존에 등록되지 못한 전문용어 및 외래어에 대한 부정확한 추출 사례들이 존재하였다. 또한 추후 연구에서는 합성어 처리 및 개선된 문장 벡터 생성 등에 대하여 진행할 예정이다.

ACKNOWLEDGMENT

이 연구는 2019년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임(10077260).

참고문헌

[1] Zameer Gulzar, A. Anny Leema, and Gerard Deepak. "Pcrs: Personalized course recommender system based on hybrid approach." Procedia Computer Science, Vol. 125, pp. 518-524, 2018.

[2] Jun Xiao, Minjuan Wang, Bingqian Jiang, Junli Li. "A personalized recommendation system with combinational algorithm for online learning." Journal of Ambient Intelligence and Humanized Computing Vol. 9, No. 3, pp. 667-677, 2018.

[3] Shuai Zhang, Lina Yao, Aixin Sun, Yi Tay. "Deep learning based recommender system: A survey and new perspectives." ACM Computing Surveys (CSUR), Vol. 52, No. 1, Article No. 5, 2019.

[4] 서봉원. "콘텐츠 추천 알고리즘의 진화." 한국 콘텐츠 진흥원 방송 & 트렌드 인사이트 5, 2016.

[5] Lalita Sharma and Anju Gera. "A survey of recommendation system: Research challenges." International Journal of Engineering Trends and Technology (IJETT) Vol. 4, No. 5, pp. 1989-1992, 2013.

[6] Mi Zhang, Jie Tang, Xuchen Zhang, Xiangyang Xued. "Addressing cold start in recommender systems: A semi-supervised co-training algorithm." Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval. pp.73-82, 2014.

[7] E. L. Park and S. Cho, "KoNLPy: Korean natural language processing in Python," in Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology, pp. 133-136, 2014.

[8] Nair, Vinod, and Geoffrey E. Hinton. "Rectified linear units improve restricted boltzmann machines." Proceedings of the 27th international conference on machine learning (ICML-10), 2010.

[9] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.