

자연어 처리와 텐서플로를 통한 언어표현 수치화 및 데이터 매칭에 대한 연구

김은진¹, 김지혜², 김치훈³, 배채은⁴, 김영종*
^{1,2,3,4,*} 숭실대학교 소프트웨어학부

e-mail: katherine0115@naver.com, rlawo0307@naver.com,
doolysmile1@naver.com, bchaeun@naver.com, youngjong@ssu.ac.kr

Research on language numerlization and data matching through natural language processing and tensorflow

Eunjin Kim¹, Jihye Kim², Chihun Kim³, Chaeun Bae⁴, Youngjong Kim*
^{1,2,3,4,*}School of Software, Soongsil University

요 약

일상생활에서 사람들은 각자 자신의 맞춤형 생활을 원한다. 특히 쇼핑이나 강의 등 직접 사용한 자의 후기에 따라 구매를 하는 경우에는 선택이 중요하다. 따라서 이 연구를 통해 머신러닝의 속성 범주화로 사용자에게 꼭 맞는 제품과 강의를 연결 할 수 있도록 한다.

1. 서론

대학생 사이에는 ‘완벽한 시간표’란 ‘좋은 강의로 채워진 시간표’를 의미한다. 학생들은 강의 계획서를 확인하고 본인에게 맞는 강의를 선택한다. 그런데 수강신청 이후에 강의 계획서가 업로드 되는 경우도 존재하며 내용의 완성도 또한 높지 않은 경우가 많다.

당번 강의계획서 '135개' 도대체 언제? - 경북대학교 인터넷 신문
www.knun.net/news/article.htm?no=15871
2012. 9. 10. - 보건복지학부는 개설강의 68개 중 15개(22%)의 강의계획서가 올라오지 않았다. 이러한 강의계획서의 부실함에 학생들은 강한 불만을 표했다.

한국일보: "수강신청한 게 언제데 강의계획서 안 나오나 ...
www.hankookilbo.com/v/377b351ea0874188a51acecf9f134932
2014. 9. 27. - 서울대 학생들이 수강신청 기간에 강의계획서를 제공하지 않는 교수가 많아 불만을 터뜨리고 있다. 하지만 대학은 이런 강의가 얼마나 되는지 파악 ...

(그림 1) 강의 계획서 업로드 문제 기사

학생들은 강의 계획서로 선택할 수 없는 경우에는 강의 평으로 강의를 선택한다. 학교에 공식적으로 업로드 되는 강의평은 강의 만족도이다.

하지만 학생들은 강의 만족도만으로는 본인이 듣고 싶은 강의를 선택할 수 없다.

학생들은 대학생 커뮤니티 홈페이지나 강의평을 따로 모아놓은 홈페이지를 통해 교수님의 수업 스타일, 시험 스타일을 확인한다. 따라서 이 연구를 통해 강의평을 분석하고 사용자의 수업 스타일에 따른 강의를 연결하는 프로그램을 개발하도록 한다.

또한, 궁극적으로는 강의뿐만 아니라 제품을 구입하거나 영화, 공연 등을 선택에 적용 될 수 있도록 확장하는 것이 목표이다.

2. 연구 내용 및 계획

○ 연구 내용

본 연구를 통해 자연어 처리로 강의평을 형태소나 구문 단위로 성분을 분석하고 TensorFlow를 이용하여 강의평을 긍정과 부정으로 분류하고 속성을 분류하여 각 속성마다 어느 정도의 값을 가지고 있는지 수치화한다.

○ 연구 목표

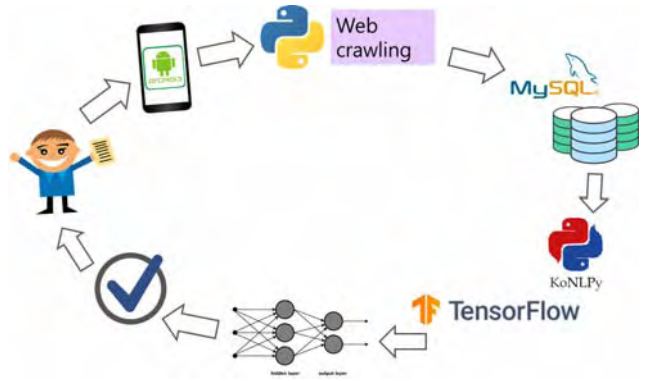
강의평은 가독성이 낮은 긴 내용들의 평가로 이루어져 있기 때문에 사용자가 원하는 조건들로 뽑아 맞춤형 강의를 찾으려 하는 것이다. 평가에서 한 가지의 속성이 아니라 여러 가지의 속성을 분석하여 범주화를 한 뒤 사용자의 속성과 비교하여 일치하는 확률이 높다면 매칭, 그렇지 않다면 보이지 않도록 하여 선택하지 않도록 한다.

또한, 소프트웨어 프로젝트, 프로그램 등록을 최소 목표로 한다.

○ 연구 예상 일정 및 범위

예상 일정	예상 연구 범위
2019년 3월	웹 크롤링 : 하나의 페이지에서 크롤링 연구
2019년 4월	- 웹 크롤링 : 다수의 페이지에서 크롤링 연구 - 자연어 처리 : 한글 및 공백을 제외한 문자 제거
2019년 5월	- 자연어 처리 : 형태소 나눔, 불용어 제거 연구 - TensorFlow : 속성 modeling 연구
2019년 6월	- TensorFlow : model과 test data 사이의 유사도 연구 - Android : UI 구성, TensorFlow와 연결하여 구현

○ 구동 예시

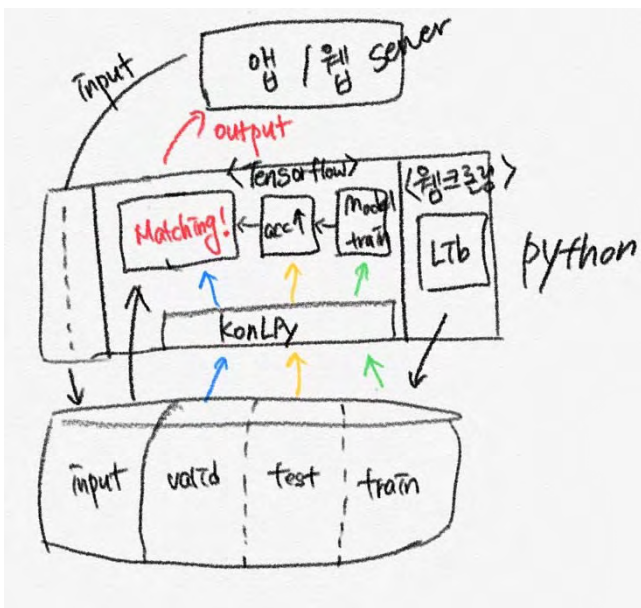


3. 프로그램 구성

○ 예상 앱 UI 구상도



○ 예상 내부 아키텍처



4. 기대 효과

본 연구를 통해 자연어 처리와 TensorFlow로 강의 평을 분석하고 속성을 나눠 수치화함으로써 사용자가 지정한 중요도에 따른 맞춤형 강의를 객관적으로 선택할 수 있도록 한다.

또한, 이러한 기능은 쇼핑 후기, 영화평, 음악평, 공연평 등 다양한 분야에 적용될 수 있으며 더불어 단순한 평점에 의존한 점수처리가 아닌 후기를 다양한 관점으로 언어 문맥에 따라 분석한다는 점에서 기존에 제공되는 평가 시스템보다 더 체계화되고 객관적인 시스템의 도입으로 이어질 것으로 기대되어 지는 바이다.

이러한 기능을 사용함으로써 사용자는 단시간에 만족도 높은 효율적인 선택을 할 수 있을 것이다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW중심대학사업의 연구결과로 수행되었음 (2018-0-00209-001)"

참고문헌

[1] 강남역까만콩, 「좋은 강의와 '대학 강의 평가」, 2016.01.30.1면, (<https://brunch.co.kr/@five0203/16>).