

데이터 사이언스 School 추천 시스템

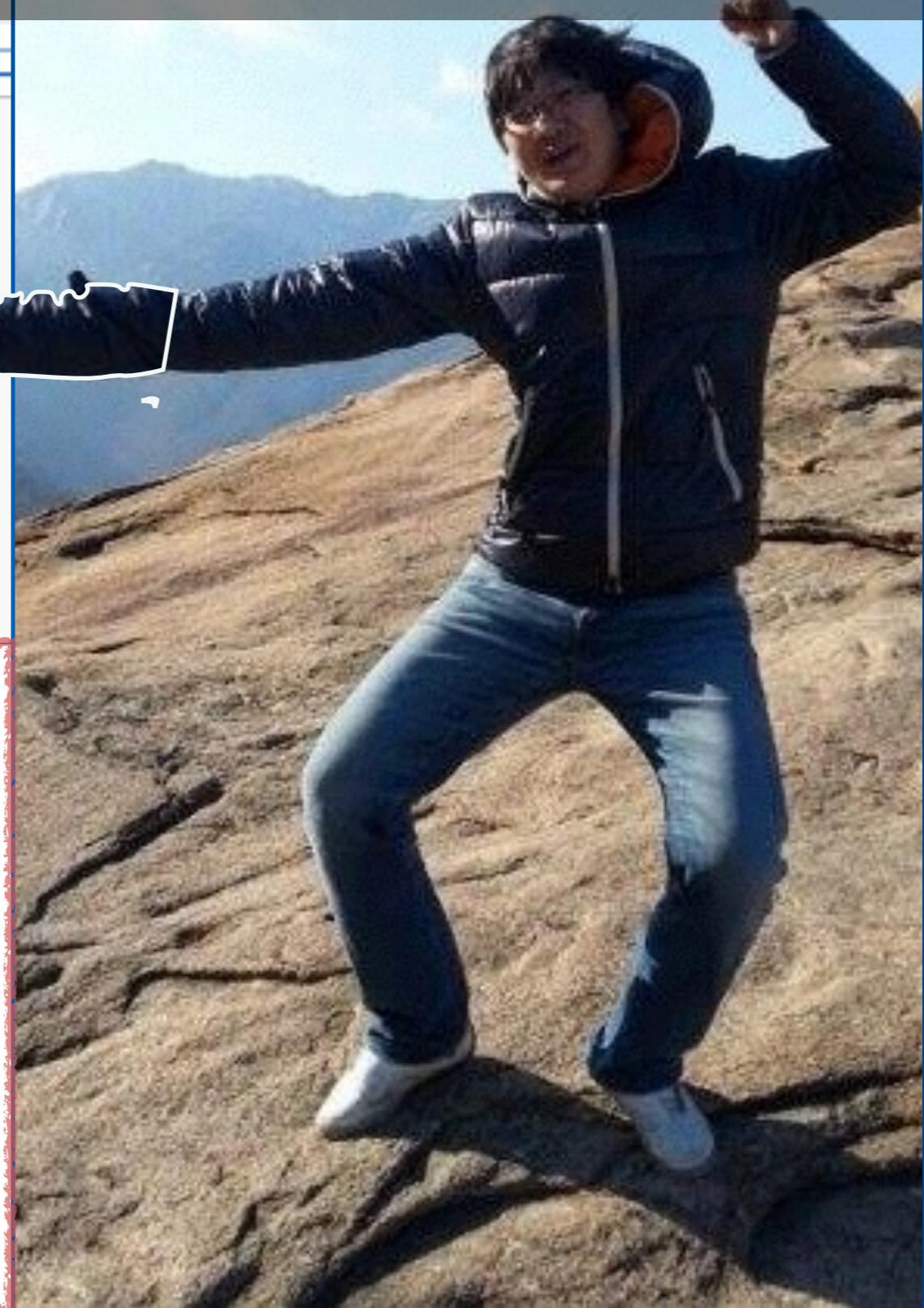
최규민(goodvc78@gmail.com)

저(최규민)는 아프리카TV에서 추천 시스템을 만들고 있습니다.

The screenshot shows the AfreecaTV website interface. At the top, there's a navigation bar with various icons and a search bar. Below it is a grid of smaller video thumbnails. A specific thumbnail of a woman is highlighted with a red border and a white hand cursor icon pointing at it. The main video player window below shows a woman with long dark hair, wearing a pink top, sitting in front of a microphone. She has a small white teddy bear with a crown on her desk. The video player includes a progress bar at the bottom showing '00:00:17 / 00:05:00'. To the right of the video player is a sidebar with a list of messages from viewers. The messages are:

- [생방]★추천업★ 페니자 개쥬마회장△(love106guy): 00:00:00 이후에 저장된 채팅입니다.
 - PC-화면하단 추천 모바일-터치후오른쪽상단 따봉/뿅/ * 가슴&컵 성드Robin 봉어리및강희! *
 - * 이그로 봉어리입니다! *
 - * 매너채팅 부탁드립니다 *
- 지네스(kimh6358): 졸아씨 배꼽보여주세요
- 음하하핫ㅋ(hamleet): 누나 누 인사 한번마누~

At the very bottom of the sidebar, there's an advertisement for 'afreecaTV AD'.



시작하기에 앞서
학습 (Learning)이란?

케이션에서 해결해야 하는 문제 영역도 탐구해야 하고 사공사 기관에서도 변화하는 것도 읽어야 하며 프야 합니다. 팀 동료의 기벽도 익혀야 하고 자기 분야의 흐름이 배우고 또 배워야 합니다. 배움의 힘으로 매일매일 쏟아지는 오래된 문제들과 새로운 문제들의 집중포화에 맞서는 것입니다.

원리 자체는 간단해 보이지만 배움이라는 것, 비판적인 사고, 창조성, 발명, 즉 마음의 능력을 확장하는 기술들은 그리 쉬운 것이 아닙니다. 가르침을 받는 것으로 될 일이 아닙니다. 배움의 능력이 필요합니다. 우리는 교사와 학생의 관계를 착각하는 경우가 많습니다. 학습은 교사가 가르치는 게 아닙니다. 학생이 배우는 것입니다. 배움은 전적으로 여러분에게 달린 것입니다.

프로그래머는 지속적으로 학습해야 합니다. 단순히 신기술을 배우는 것만은 아닙니다. 팀 환경의 변화, 협업 기법, 그리고 자기 분야에서의 능력이 변화하는 것도 읽어야 하며 프로젝트 자체가 완성되는 정도, 퀴즈체인 할인 코드, 소프트웨어 개발의 리더 역할 등은 그 배움의 힘으로 매일매일 즐아지는 오래된 문제들과 새로운 문제들입니다.

것입니다.

여러분이 주체이자 주인공입니다.

원리 자체는 간단해 보이지만 배움이라면 배움의 능력, 즉 마음의 능력을 확장하는 기술들은 그리 쉬운 것이 아닙니다. 가르침을 받는 것으로 될 일이 아닙니다. 배움의 능력이 필요합니다. 우리는 교사와 학생의 관계를 착각하는 경우가 많습니다. 학습은 교사가 가르치는 게 아닙니다. 학생이 배우는 것입니다. 배움은 전적으로 여러분에게 달린 것입니다.

이 책『실용주의 사고와 학습』을 통해 여러분의 학습 능력이 더욱 향상되고 더 실용적으로 생각할 수 있게 되기를 바랍니다.

앤디 헌트 저(박영록옮김) “실용주의 사고와 학습” 발췌, 위키북스

애자의 바버로이 무엇이까요?

■ 단계 2: 특징 선택 단계

앞에서 이야기한 것처럼 패턴인식 시스템의 성능에 결정적인 영향을 미치는 단계다. 대상 패턴에 대한 충분한 사전 분석을 통하여 어떠한 특징을 선택할 것인지를 이 단계에서 결정하여야 한다.

■ 단계 3: 모델 선택 단계

패턴인식을 위한 여러 접근법 중에서 어느 모델을 어떠한 알고리즘을 이용하여 어떻게 구성할 것인가를 결정하는 단계다. 특징 선택과 마찬가지로 사전에 특징에 관한 경험적인 지식이 필요하며, 각 접근법에 따라 모델에 필요한 각종 파라미터를 설정해야 한다.

■ 단계 4: 학습 단계

수집된 데이터에서 추출한 특징 집합을 활용하여 선택한 모델이 있을 때 학습을 통해 비어 있는 블랭크(blank) 모델을 완전한 모델로 만드는 단계다. 학습은 그 방법에 따라서 크게 교사학습, 비교사 학습 그리고 강화 학습으로 나뉜다.

■ 단계 5: 인식 단계

패턴인식의 마지막 단계로, 임의의 특징이 주어지면, 이 특징이 속한 클래스 혹은 카테고리를 결정하는 단계다.

브런치 작가
추천해 보기

유사도와
군집화

영화
별점 예측

word2vec과
추천의 만남

$$aX + b$$

최적해 $[a, b] = [2, 3]$

Y

추천 시스템 구현

여러분이 풀고자하는 문제

X

$$2X + 3$$

Y

여러분의 추천 시스템 구현

Overview

강의 목차

ch1. 브런치 작가추천해 보기

ch2. 유사도와 군집화 알고리즘

ch3. 영화 별점 예측하기

ch4. Word2Vec과 추천 시스템 만남

ch5. 넷플릭스 추천의 힘

ch1. 브런치 작가 추천해보기

광고로 알아보는
나의 관심사

최근 나의 추천 사례

페이스북 광고로
나의 관심사 분석

브런치 작가
추천 시스템 만들기

데이터를 모으자
좋아하는 작가 분석 및 추천

ch2. 유사도와 군집화

유사도 알고리즘

cosine, euclidean,
jaccard, correlation

군집화 알고리즘

K-Means,
Hierarchical Clustering



ch3. 영화 별점 예측

User Based CF
영화 별점 예측

MovieLens
데이터셋 활용

유저	별점	유사도	weighted	none
A	Nan	0.61	-	
B	4	0.91	3.64	4
C	3	0.63	1.89	3
D	4	0.91	3.64	4
합계		9.17	11	
예측별점		3.74	3.67	

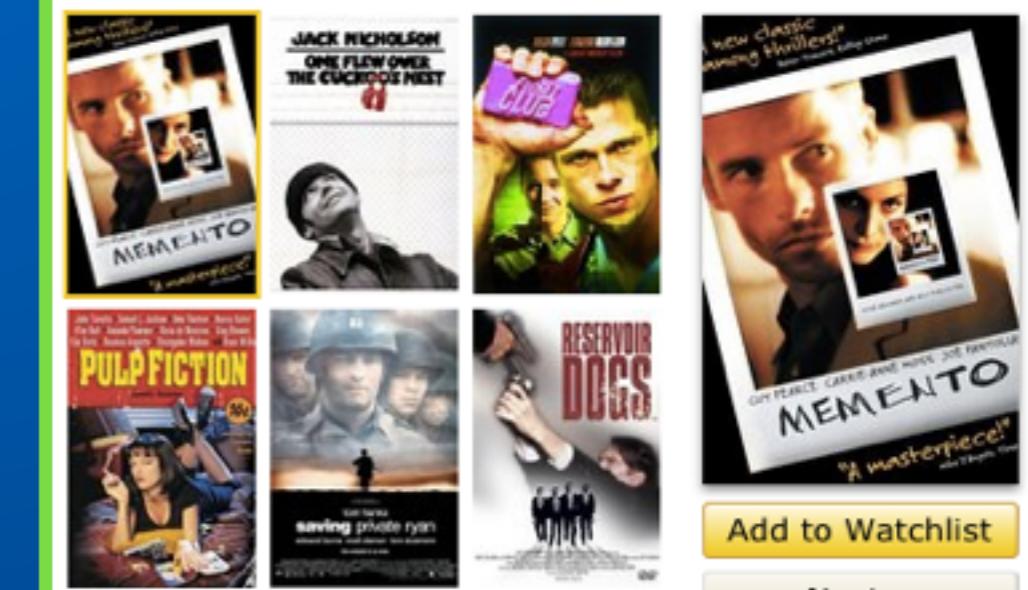
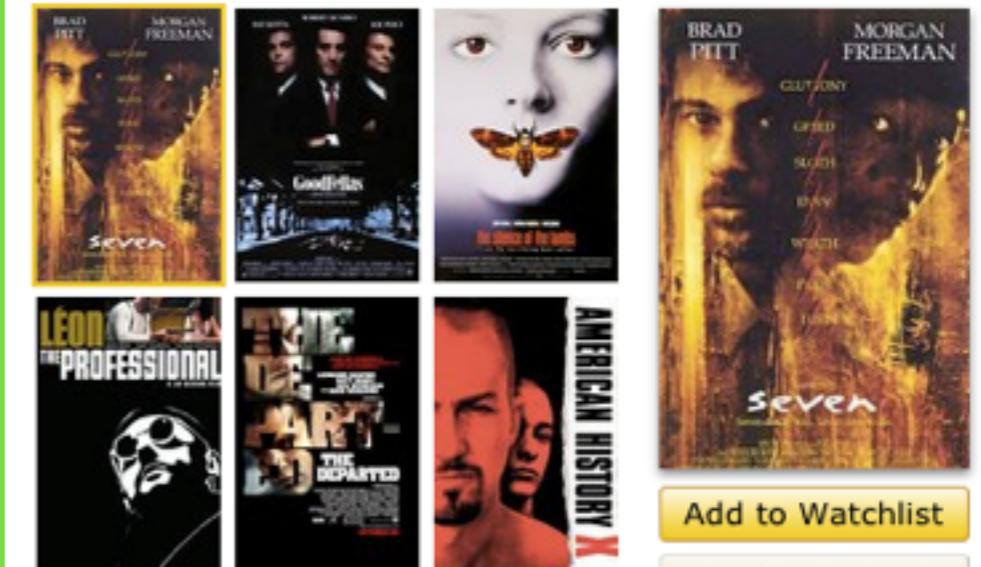
ch4. Word2Vec이 추천 시 스템을 만났을때

Word2Vec에 대
하여

$$\arg \max_{\theta} \prod_{(w,c) \in D} p(c|w; \theta)$$

Movie2Vec
만들기

People who liked this also liked...



이론 설명
/토론

1.5H

데이터의
이해

0.5H

직접
구현하기

1H



실습 환경 셋팅

실습환경

- 노트북 필수
- Python 3.4.3 → 2.7
- Jupyter Notebook(IPython)
 - Interactive Web IDE 환경
- 주요 사용 패키지
 - Pandas, Numpy, Scipy, Scikit-Learn, selenium, seaborn,
- 파이썬을 활용한 데이터 분석 (2 ~ 9장) 참고
 - 웨스 메카니 지음, 김영근 옮김
- 실습 자료 저장소(git) : <https://github.com/goodvc78/fc-recommender-school>

주로 사용하는 Python 패키지

- Pandas : Pannel Data Analysis Package
 - 참고: <http://www.slideshare.net/maikroeder/pandas-16424935>
- Numpy : Numeric Python Package
- Scipy : Science Python Package
- Scikit-Learn : Machine Learning Python Package
- matplotlib : 2d plotting library

주로 사용하는 Python 패키지

- Gensim : topic modeling package(for word2vec)
- networkX : network graph
- BS4 : html parsing
- urllib : url request
- selenium : UX Test Automation for web application
- seaborn : statistical data visualization library

최규민

- 이메일 : goodvc78@gmail.com
- Facebook : <https://www.facebook.com/goodvc78>
- 블로그 : <https://brunch.co.kr/@goodvc78>(현)
<http://goodvc.tistory.com/> (구)

감사합니다.