[문서 제목]

[문서 부제]

중간보고

\* 발표자료 2분(ppt 2~3장)

\* 중간보고서 별도

\* 본인 깃허브 링크

\* 프로그램 소스

\* 모든 소스에 라이선스 명시

1. 주제 :

2. 필요성 및 기대효과 :

3. 기존 연구 :

- 사용 오픈소스 소프트웨어 설명(텐서플루우, CNN, RNN, LSTM, GAN,...)

- 현재 상위 방법들 설명

4.  제안 방법 :

- 전체 구성도 혹은 흐름도(플로우 차트)

- 제안 방법 상세 설계(단계별 설계)

5. 실험 및 결과(비교실험) :

- 데이터베이스 설명( 출처, 갯수, 구성, 시기,...)

- 사용환경(운영체제, 사용언어, 라이브러리 버전, 하드웨어)

- 현재 상위 방법들과 비교

6. 고찰(토론) :  실패 케이스와 그 이유....

7. 참고 문헌 :

[1] 저자명, "제목", 출처(출판사 등), 판차, 년도.

8 느낀점 :

- 본인 깃허브 사용 히스토리 첨부

\* 기존연구에서 소개된 내용을 실험에서 비교하기

\* 참고한 자료가 있는 곳 문장 끝에 참고 문헌에 기록된 번호를 기입하기

----------

\* 제안 방법에 들어갈 내용

1. 문제 파악 (주제 선정)

2. 데이터베이스 다운로드 (자료 수집)

3. 필요 데이터 선별(불필요 데이터 제외)

4. 예측(prediction) 모델에 맞게 데이터 변형(데이터 가공)

\* 실험및 결과에 들어갈 내용

5. 모델 학습(data train based on model)

- 실험 데이터에 대한 설명(자료형태, 구조)

- 개발환경(H/W 스펙, SW 버전)

6. 모델 테스트(data test based on trainning model)

7. 모델 개선(Improve model using my unique idea)

고유한 나만의 아이디어는 가. 전혀 새로운 접근  나. 기존의 방법의 조합 등의 형태이다.

내용

[1. 주제 2](#_Toc534744992)

[2. 필요성 및 기대효과 2](#_Toc534744993)

[3. 기존 연구 2](#_Toc534744994)

[4. 제안방법 2](#_Toc534744995)

[- 문제 파악 2](#_Toc534744996)

[- 데이터베이스 다운로드 2](#_Toc534744997)

[- 필요 데이터 선별 2](#_Toc534744998)

[- 예측 모델에 맞게 데이터 변형 2](#_Toc534744999)

[5. 실험 및 결과 2](#_Toc534745000)

[- 실험 데이터에 대한 설명(자료형태, 구조) 2](#_Toc534745001)

[- 개발환경(H/W 스펙, SW 버전) 2](#_Toc534745002)

[- 모델 테스트 2](#_Toc534745003)

[- 모델 개선 2](#_Toc534745004)

[6. 고촐(토론) :실패 케이스와 그 이유 2](#_Toc534745005)

[7. 참고 문헌 2](#_Toc534745006)

[8. 느낀점 2](#_Toc534745007)

# 1. 주제

* 영화의 정보들을 모아둔 데이터를 사용하여 사용자에게 영화를 추천해주는 프로그램 구현

# 2. 필요성 및 기대효과

온라인 컨텐츠 시장이 성장함에 따라 소비자들은 더욱 많은 선택권을 가지게 되었다. 하지만 그만큼 모든 컨텐츠를 소모하기엔 시간적 여유가 따르지 않기 때문에 소비자들은 제한된 시간을 효율적으로 소모하기 위해 제한적으로 컨텐츠를 고를 수 밖에 없다. 대부분의 멀티미디어 서비스 웹사이트들은 이런 소비심리를 전략적으로 이용하기 위해 대부분 컨텐츠 추천 알고리즘을 사용하고 있다.

최근 가장 높은 성장률을 보이고 있는 넷플릭스(Netflix)는 대표적인 멀티미디어 컨텐츠 서비스 사이트이다. 넷플릭스는 강력한 컨텐츠 추천알고리즘을 사용하여 사용자(User)가 시청한 영상물들을 분석하여 트렌드를 파악하고, 취향에 맞는 컨텐츠를 추천해준다. 또는 다른 유저들이 많이 즐긴 컨텐츠들을 리스트로 상단에 노출시키는 마케팅을 하기도 한다.

이런 컨텐츠 추천 알고리즘은 소비산업에서 가장 필수적이면서 기본적인 마케팅 기법이 되었다. 미디어 산업뿐만 아니라 쇼핑, 스포츠, 부동산등 여러 분야에서 활발히 사용되고 있고 소비산업 뿐만 아니라 사용자들의 니즈를 분석하여 새로운 데이터를 수집할 수 있다는 점에서 다양하게 사용될 수 있다. 앞으로 이런 추천 알고리즘을 개량시키고 발전시킴으로써 궁극적으로 소비자의 마음을 읽어 완벽하게 필요로 하는 요소들만 추천하는 인공지능 분야에도 사용될 수 있다.

# 3. 기존 연구

# ìí ì¶ì² ìê³ ë¦¬ì¦ì ëí ì´ë¯¸ì§ ê²ìê²°ê³¼

<http://blog.kthdaisy.com:8080/recommendation_system_kthdaisy/>

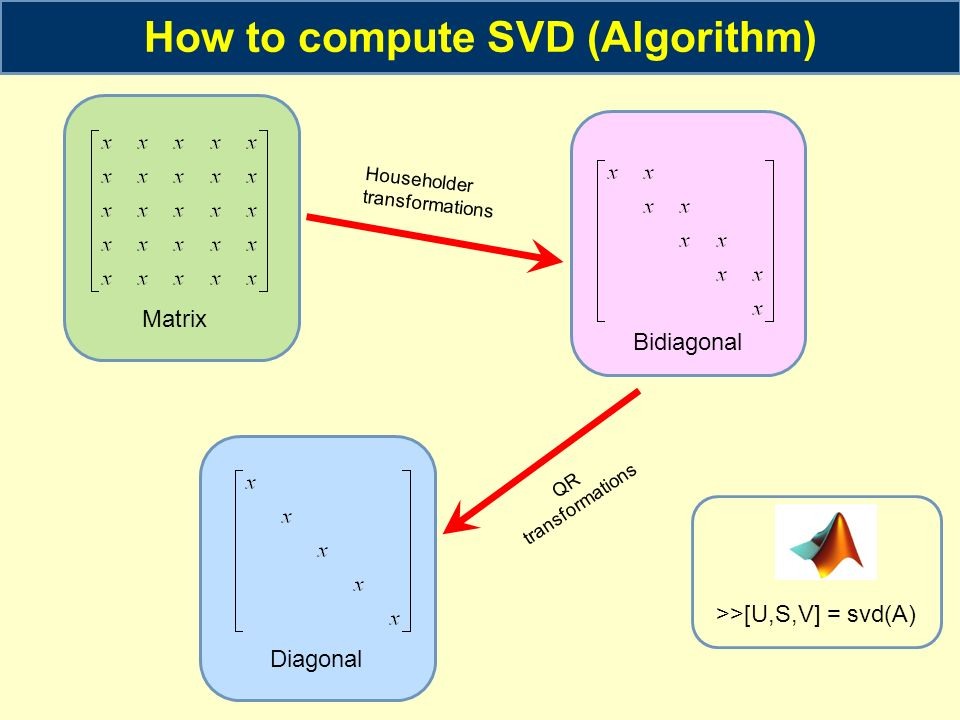
현재 가장 활발히 사용되고 있는 추천 알고리즘은 협업 필터링(Collaborative filtering)이다. 협업 필터링은 사용자들로부터 정보를 받아 사용자들의 관심사들을 자동적으로 예측하여 추천해주는 필터링 기법이다. 협업 필터링은 크게 유저 기반 협업 필터링(User-based-CF), 아이템 기반 필터링(Item-based-CF)로 나누어지고, 유저 기반 협업 필터링은 다시 능동적, 수동적 필터링으로 분류 된다. 이런 필터링 알고리즘들은 데이터를 분석하여 유사도 계산을 통해 추천하게 된다.

유저 기반 필터링의 소개와 장점과 단점

능동적, 수동적

아이템 기반 필터링의 소개와 장점과 단점

협업 필터링의 단점과 그것을 개선시킬 SVD



SVD소개

<https://slideplayer.com/slide/8816430/>

# 4. 제안방법

- 전체 구성도 혹은 흐름도(플로우 차트)

- 제안 방법 상세 설계(단계별 설계)

스크린샷이(가) 표시된 사진

자동 생성된 설명

## - 문제 파악

협업 필터링의 단점에서 문제점을 파악한다.

## - 데이터베이스 다운로드

kaggle에서 제공하는 데이터셋 다운로드

## - 필요 데이터 선별

데이터셋을 변경하자

## - 예측 모델에 맞게 데이터 변형

# 5. 실험 및 결과

## - 실험 데이터에 대한 설명(자료형태, 구조)

주피터 노트북 실행 화면

## - 개발환경(H/W 스펙, SW 버전)

H/W

|  |  |
| --- | --- |
| 하드웨어 | 스펙 |
| CPU | i7-8770K |
| GPU | GTX-1060C |
| RAM | 8G |
| STORAGE | SSD 250GB, HDD 2TB |

SW

|  |  |
| --- | --- |
| 소프트웨어 | 버전 |
| 윈도우10 |  |
| Anaconda Navigator |  |
| Jupyter Notebook |  |
| Python |  |

## - 모델 테스트

## - 모델 개선

# 6. 고찰(토론) :실패 케이스와 그 이유

# 7. 참고 문헌

# 8. 느낀점