

Techtonic 2018

-
Thu . Nov 15
-
SAMSUNG SDS Tower
West Campus B1F
Magellan Hall / Pascal Hall



SAMSUNG SDS
Realize your vision

Deep Auto-encoder로 추천 알고리즘 만들기

삼성SDS 주성훈 프로



- 콘텐츠 추천
- 딥러닝을 기반 콘텐츠 추천
- 실제 추천 사례
- Self Q & A

Deep Auto-encoder로 추천 알고리즘 만들기

컨텐츠 추천

1998년 11월 15일 동아일보 TV가이드

오늘의 TV 서울·인천·강원

○오전 ●오후 ●밤 ◻스페셜 ◻음성다중 ●문의=KBS 781-1800 MBC 780-0011 SBS 369-1090 ITV 032-830-0900
●이 프로그램은 사정에 따라 변경될 수 있습니다. ●프로그램 열의 숫자는 G코드 (문의처 02-737-8088)

	KBS 1	KBS 2	MBC	SBS	iTV
6	00 뉴스 4300270 10 농업도 결연이다 27712486	00 세계는 지금 8334047 45 동물의 왕국 6869950	00 TV 메디컬 센터 68318 00 뉴스 7317848 20 재미있는 만남 사람 20583047	00 출발 모닝와이드 일요일 특집 <1, 2, 3부> 3551028	00 아침뉴스 7236757 10 굿모닝 골프박사 89951047 00 뉴스 5428689 10 문화살롱 64029347
7	00 사랑의 가족 14912 30 시청자 의견... 33047	15 신 한국기행 16173115 <천재통의 울산> 00 디즈니 만화동산-전사 클리앙 3282283 50 시네마 데이트 28655134	00 뉴스 7317848 20 재미있는 만남 사람 20583047	05 김동길의 선데이... 80692202 55 장사특집 다큐-하말 라아의 사나이4 87637554	00 건강 25시 56080 30 일요만화광국 7234689
8	00 뉴스 6043009 15 전채전단 72906931	00 디즈니 만화동산-전사 클리앙 3282283 50 시네마 데이트 28655134	00 디즈니 만화동산-전사 클리앙 3282283 50 시네마 데이트 28655134	05 김동길의 선데이... 80692202 55 장사특집 다큐-하말 라아의 사나이4 87637554	00 건강 25시 56080 30 일요만화광국 7234689
9	00 9국악 한마당 <금강산 타령, 함경도 풍구소리 등> 97318	40 도전 지구탐험대 <김명덕의 일류체험-독일 트렘플린, 구자미의 문화체험-인 NAL라다크> 64866912	00 9일요아침드라마-사랑밖에 난 몰라 <적과의 동거> 82486	40 LA아리랑 57410060	50 뉴스 27533757
10	00 대하 드라마 왕과 비 <44, 45회> 70844	40 도전 지구탐험대 <김명덕의 일류체험-독일 트렘플린, 구자미의 문화체험-인 NAL라다크> 64866912	00 사람의 스튜디오 진행:임성훈 양혜영 24399	50 좋은 친구들 <비교체험 극과극-SBS방송국과 동국대 방송국을 비교> 진행:남희석 박수홍 55091047	00 사랑의 뮤직쇼에 NAVER 307776
11	30 다큐멘터리 극장 69825	00 샬리리엔 일요일이 좋다 <서울은행 편> 5277979	00 전일일기 카본:이은정 연출:최용원 37863	00 뉴스 3017979 10 최고의 밤상 76569757	30 안보 지그재그 <강화도2> 6469592
12	00 뉴스 6344554 10 전국 노래자랑 8220689	50 쇼 행운을 잡아라 83667663	00 뉴스 6362950 10 출발 비디오 여행 19416912	00 뉴스 3017979 10 최고의 밤상 76569757	20 뉴스 35551573 30 아프리카의 동물들 85080
1	10 2020년 우주의 원더 키드 78793573	00 뉴스 8149450 10 강호홍운록 86235738	00 앙코르 남자셋 여자 셋 3756047	00 서세원의 좋은 세상 만들기 11738	00 불우청소년을 위한 박찬호 자선 디너쇼 <1, 2부> 49863
2	00 비디오 챔피언 <98-99 프로농구 대우:현대> 3854660 55 98-99 프로농구 <대우:현대> 13160554	00 화인 메일을 벗겨라 <에> 5688825 50 주말 연속극-종이 락 <4, 5회> 99520252 극본:조소혜 연출:김종창	20 김국진 김용만의 21세기 위원회에 43475028	00 기쁜 우리 토요일에 50202	00 98프로농구 <삼성:나래> 941783
3	40 비디오 추적 놀라운 TV 28305318	00 TV쇼 진품명품 6792283 55 긴급구조 119 36692283	10 주말연속극-사람과 청춘 2000979 10 레니게이트 <아내를 쓴 수사관> 36678405	00 주말극장-흐린날에 쓴 편지 <4, 5회> 70318 극본:김운경 연출:이종수	00 뉴스 7824931 10 다크스카이 54590134
4	00 뉴스 3537757 10 세상처럼 아파와... 69791134 25 힘내세요 사장님 57617776 15 재미있는 동물의 세계 81847689	00 TV쇼 진품명품 6792283 55 긴급구조 119 36692283	00 뉴스 3522825 10 10대 세상 내일이 보려고 33524318	00 뉴스 5337848 10 인기가요 21991776	00 뉴스 7824931 10 다크스카이 54590134
5	00 9월엔 음악회 출연:김진원 박연미 류재광 양희은 신애령 18689	50 9일요일은 즐거워 <진행:임백천 강성연 강호동 김현주> 19096660	00 9월엔 TV와 나의 3558399 50 일요일 일요일 밤에 <기획:이경실 김국진 신동엽> 75351467	00 홍수권의 호기심 천국 24202 00 그것이 알고싶다 <기획:이경실 김국진 신동엽> 76689	00 뉴스 7824931 10 다크스카이 54590134
6	00 일요스페셜 <만행> 94009	00 주말 연속극-종이 락 <4, 5회> 72863 극본:조소혜 연출:김종창	00 주말연속극-사람과 청춘 2000979 10 레니게이트 <아내를 쓴 수사관> 36678405	00 뉴스 5337848 10 인기가요 21991776	00 뉴스 7824931 10 다크스카이 54590134
7	00 뉴스 9 4901115 35 스포츠 뉴스 64548115 45 대하드라마-왕과 비 54228405	00 특별기획 10부작 북한영화 <림격전-마지막화> 1663450 50 뉴스 99799592	00 뉴스데스크 4996283 35 스포츠 뉴스 64533283 45 시사매거진 2580 54213573	00 뉴스 5337848 10 인기가요 21991776	00 뉴스 7824931 10 다크스카이 54590134
8	35 경제전망대 7980573	00 스포츠뉴스 8132252 10 일요 베스트 <가을날의 동화> 63577979	35 다큐멘터리 상공시대 <배상연 국순단 회장> 84704080	00 8 뉴스 5722844 35 스포츠가 좋아요 28413641 50 주말극장-흐린날에 쓴 편지 3879912	00 퀘즈쇼 베스트5 52028 30 복결판개 박찬호가 말하는 나의 15승 이야기 414370
9	05 9월명화극장 <귀여운 벨리> 91891047	15 정범규의 세상읽기 출연:안성기 나성린 36749221	25 9일요심야극장 <120일간의 동거> 감독:로드 다니엘 주연:조 페시 빈센트 가드너 어 48357009	50 장사특집 대작-패 어게임 감독:앤드루 사임스 주연:신디 크로퍼드 윌리엄 볼드윈 43029757	00 뉴스 3128202 15 윤상의 이밤을 사랑해<오락실> 28313738
10	45 TV해방 <내일을 여는 작가 (3)-구효서 편> 9209535	10 재시라의 세레나 데 80084644	25 9일요심야극장 <120일간의 동거> 감독:로드 다니엘 주연:조 페시 빈센트 가드너 어 48357009	30 장사특집 다큐멘터리-한국의 패류 <2부> 54370 30 메디컬드라마 ER * 1-30 나이트라인 55719 2924582	00 뉴스 3128202 15 윤상의 이밤을 사랑해<오락실> 28313738

▶▶▶ 700 유대권인사이드 -5553

2018년 10월 9일 오후 14시

아프리카TV

HOT 게임 스포츠 보.라 엔터 전체

전체 방송

[생] MLB 메이저리그 포스트시즌 실시간★LA다저스 애크랜더플로라... 우유 630 명 시청중	[생] 헌터X헌터 게이미펜X선거 판X압축대륙 ㅇㅎㅎ 애니메이션... 어나더TV 423 명 시청중	[생] 리니지m 정개철 라스타바드1업 정개철 1,115 명 시청중
[생] 개박이X싸패 현재 그랜드마스터 랭커 얼마안남았다.. 배틀그라... 개박이싸패이 5,845 명 시청중	[생] 카비 리니지 라바1업 총사 1등 엔틱카비? 영웅장비 먹올거... NC카비 551 명 시청중	[생] 경제vs안겨님 스타연습 lol [G]경제 409 명 시청중
[생] 토마토X단아냥 서면 데이뷰 Fresh토마토 4,452 명 시청중	[생] 2렘다이크는 멈추지 않습니다 해강고 드레이븐 lol 해강고강고 550 명 시청중	[생] 낭만고양이 조공데이 BJ낭만고양이 393 명 시청중
[생] 스타 엠보성 스타bj참교육 2단계 이영환 A-엠보성!! 4,190 명 시청중	[생] 해외선물-주식-리스크관리의 중요성!! 차원이 다른 매매 해선킹 541 명 시청중	[생] 낭만고양이 조공데이 BJ낭만고양이 393 명 시청중
[생] 최고음질 벨론 음악방송 BJ심장 발라드 음방 뮤직 노래방 인기... 심장 2,585 명 시청중	[생] H너 ♥나루도 ♥정주행♥24시애니실시간퀴즈#헌터#코난#... H너 ♥ 521 명 시청중	[생] 한정판특집공포레전드 드는치지마세요★공포라디오 쌤... 쌤우이™ 390 명 시청중
[생] 챌린저가 브론즈랑 듀오하면??? 강만식 LOL롤 강만식 2,124 명 시청중	[생] 리니지 원큐 단독 93렘 ㅋㅋ 원큐 501 명 시청중	[생] 자유령 실골 헌지인 모여라 흥이 캐리한다 롤 lol 승욱 382 명 시청중
[생] 오메김 안녕하세요^^ 오메김승현2 1,857 명 시청중	[생] 멜론음악 최신노래/인기가요/실시간차트/TOP100 락선 466 명 시청중	[생] 황보관 VS 황명보 코인팜API 사용법 (초본) 방송국공지사항 업... 솔트강 377 명 시청중
[생] 돈쓰는 날 기혼이 결혼식장 밖 프랭만두모빌 개좌파괴 용노님 1,794 명 시청중	[생] 알로하 현 스쿼드랭킹1위 266 8점 65승30탑3패 칸칸x재경x이... Alo 조경훈 461 명 시청중	[생] 배틀그라운드 싸패X개박이 듀오 품종 돌아왔나? BJ싸패 348 명 시청중
[생] 갑학수 단아냥 배웅하기 ^^ 세사반기차 부산역가기 파일병김학수 1,744 명 시청중	[생] 토끼예나 7킬치킨시 1500개! 배틀그라운드 토끼예나 459 명 시청중	[생] 연봉 2억 도전 회 30만원치 낚시 세계 최강자 분홍송어 분홍송어 340 명 시청중
[생] 리니지, 리니지M = 막부 = 투사 [막부] 1,448 명 시청중	[생] 해외선물 파생1번지 백전불패 절대기법 * 국내선물/비... 파생1번지 443 명 시청중	[생] 켄세이박 초고수들 다 이겨드림 스타 랠루 박민철+ + 313 명 시청중
		[생] 갈날비 도란검 진 이견 대혁명 임LOL 종탁이 679 명 시청중
		[생] 일본인@ 계룡대 군대체험하러 왔습니다!! 필!승! 유이홍 ♥ 663 명 시청중
		[생] 주식#해외선물 봉 매매 일퍼님!! <기준봉/중심/패턴/지지저... 일퍼님 642 명 시청중
		[생] 금일도 찾팅가즈아~ 하루하루 최선을 알파조7 634 명 시청중

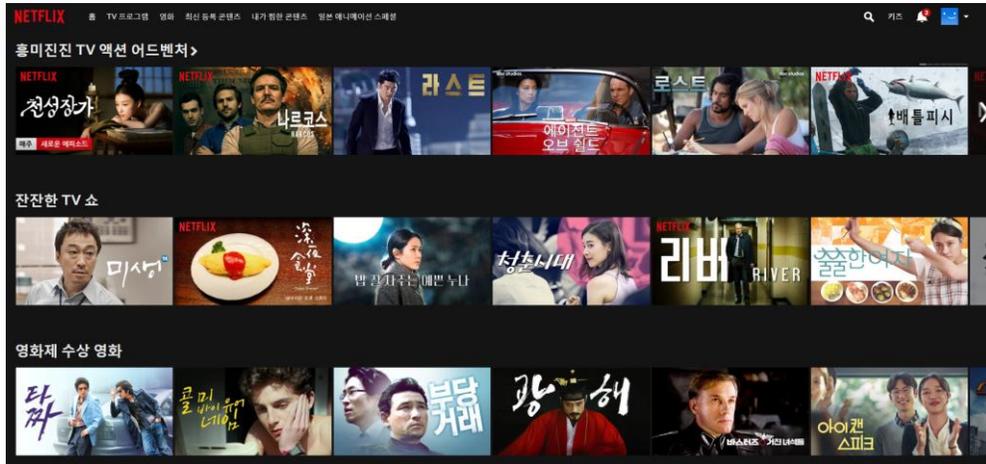
※ Source : <http://www.afreecatv.com/>

1인 미디어, 변화하는 미디어 플랫폼 시대

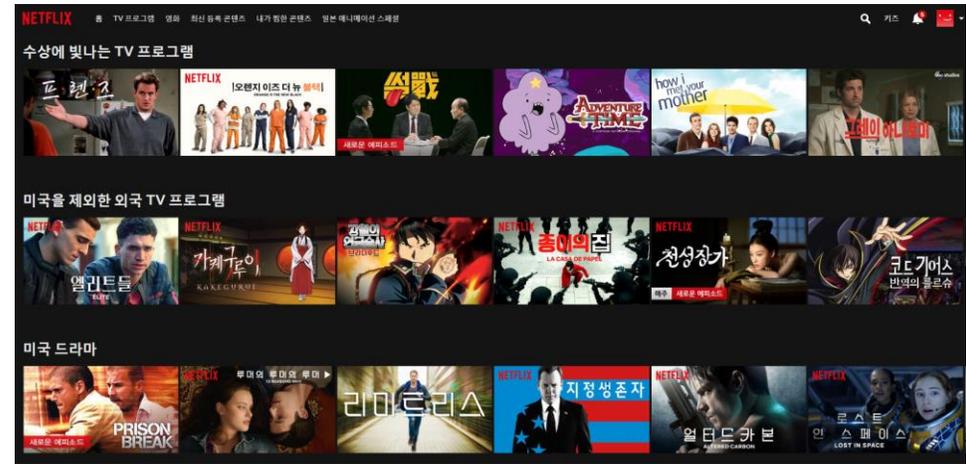


컨텐츠의 ‘개인화’

같은 플랫폼에서도 사용자 별로 서로 다른 컨텐츠가 노출



〈주성훈 님의 넷플릭스 초기화면〉



〈주성욱 님의 넷플릭스 초기화면〉

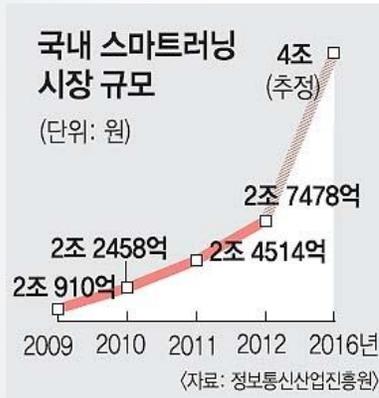
컨텐츠 추천의 필요성

- 두 시간짜리 영화를 고르기 위해 한 시간 동안 영화 리뷰를 검색하는 시간낭비
- 넷플릭스 총 시청 시간의 80%는 개인화 추천에 의한 것
- 구글 뉴스 클릭 중 38%가 구글이 추천한 링크
- 아마존 구매 중 35%가 아마존이 추천한 제품

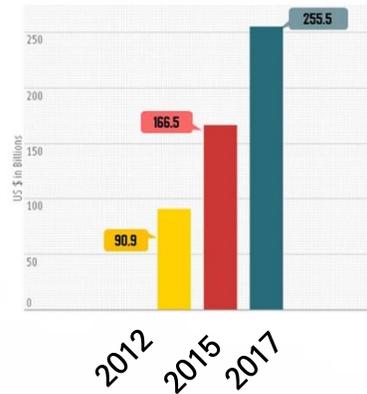
SDS는 왜 콘텐츠 추천을 연구할까?

Smart Learning/E-Learning, 광고, Smart TV 등 기업에서 활용하는 콘텐츠의 양이 지속적으로 증가

스마트 러닝



글로벌 E-learning 시장 규모



- 국내외 E-Learning 시장 규모 증가
- 다양한 분야의 교육 콘텐츠

지능형 리테일 솔루션 Samsung Nexshop



- 동영상 광고
- 상품 정보

삼성 스마트 TV



- VOD
- 방송 편성표

추천과 빅데이터 분석

효과적인 추천을 위해서는 다양한 데이터의 종합적 분석이 필수



사용자 정보



시청 시간



컨텐츠 정보



검색 내역



구매 내역



리뷰



https://www.imdb.com/title/tt1727824/?ref_=nv_sr_1



★ 3/10
The trailer looked great, the film is not
elainetaft 12 October 2017

I am ver TV show great an this, but finish, I \ whatsoe just gon
135 out of 0
Permalink

★ 4/10
Best PR team of any movie in a long time.
modulamax 15 October 2017

First I wanna say great family movie, bring the kids... So they can laugh along with you, I don't want to spoil the immense amount of nothing that happens in this movie, or how you'll grow bored of waiting for "It" to get scary, but I do want to know how much they were paying people walking into the screenings giving this movie a 10 star rating, or how large is the pr team? I laugh when I see red balloons now. Thought I was going to see a horror movie, got a comedy.. Talk about a let down.

144 out of 267 found this helpful. Was this review helpful? Sign in to vote.
Permalink

리뷰 출처: IMDB



Brightics AI

제조에서 마케팅까지 바로 적용 가능한
시 기반 빅데이터 분석 플랫폼

Deep Auto-encoder로 추천 알고리즘 만들기

딥러닝 기반 콘텐츠 추천

컨텐츠 기반 추천

과거에 소비했던 콘텐츠와 유사한 콘텐츠를 추천



시청한 영화에 대한 정보



- 감독
- 출연 배우
- 영화의 배경
- 영화의 원작
- 영화의 소재

...



유사한 영화 추천



(타짜, 내부자들, 관상) 처럼 백윤식 출연
(타짜, 내부자들) 처럼 조승우 출연
(관상) 처럼 역학 소재

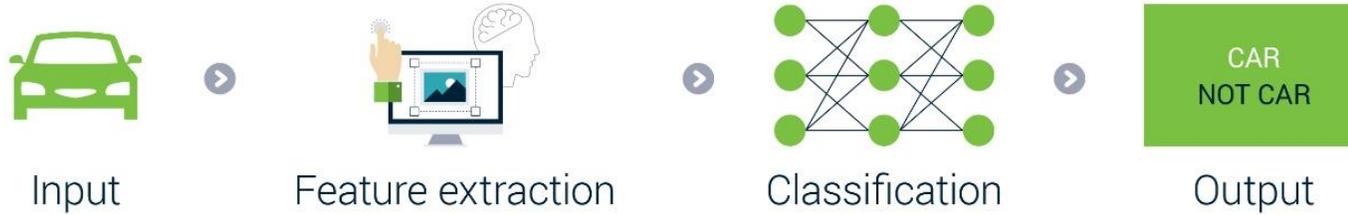


(타짜) 처럼 최동훈 감독 작품
(타짜, 관상) 처럼 김혜수 출연
(관상) 처럼 이정재 출연
(타짜, 내부자들) 처럼 범죄 영화

영화 포스터 출처: 영화진흥위원회 영화정보 DB (<http://www.kobis.or.kr/kobis/business/mast/mvie/searchMovieList.do>)

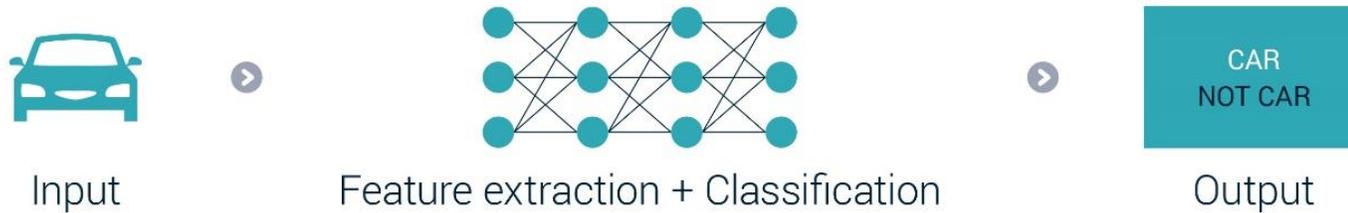
추천에서의 딥러닝 활용

Machine Learning



컨텐츠의 특징을 일일이 정의

Deep Learning



방대한 양의 컨텐츠로부터
특징을 직접 추출

- 대용량 컨텐츠의 특징을 추출 할 때 유리
- 비선형적이므로 사용자와 컨텐츠 간의 복잡한 관계를 포착할 수 있음
- 충분히 많은 데이터를 학습했을 경우 새로운 데이터에 대한 유연한 처리가 가능

딤러닝 기반 추천 – 콘텐츠 유사도 예측

영화를 딤러닝으로 만든 좌표공간에 표현한다면...



선호도 기반 추천

취향이 유사한 사용자들이 좋아하는 컨텐츠들을 추천



협업 필터링(Collaborative Filtering)

Matrix Factorization 을 이용하여 사용자의 콘텐츠 선호도 예측

사용자 x 콘텐츠 matrix

	5	0	4	0	3	3
	0	4	0	0	3	0
	3	0	0	3	0	0
	0	4	0	5	0	1

시청 (평점) 미시청

≈

Feature N=2

	0.49	2.09
	1.42	1.43
	0.82	1.08
	1.88	0.85

X

Feature N=2

0.74	1.58	0.51	2.12	0.86	-0.12
2.21	1.22	1.79	1.16	1.23	1.46

=

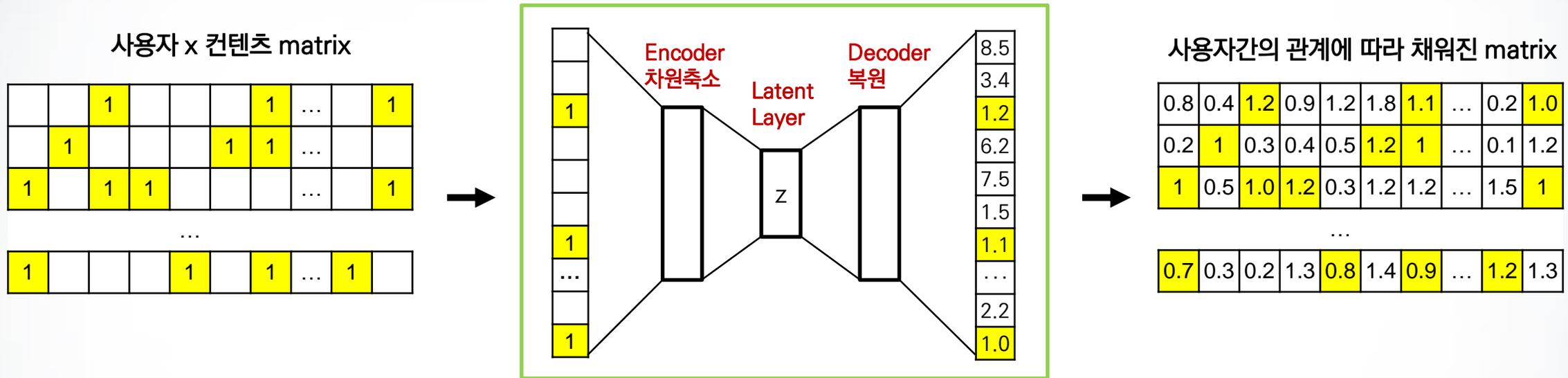
4.98	3.32	3.99	3.47	3.00	2.98
4.21	3.99	3.29	4.69	2.99	1.91
2.99	2.61	2.35	2.99	2.04	1.48
3.26	3.99	2.47	4.98	2.66	1.01

Matrix Completion
평점 예측



딥러닝을 활용한 협업 필터링 – Auto-encoder 기반

Auto-encoder의 차원 축소-복구 과정을 이용한 콘텐츠 추천



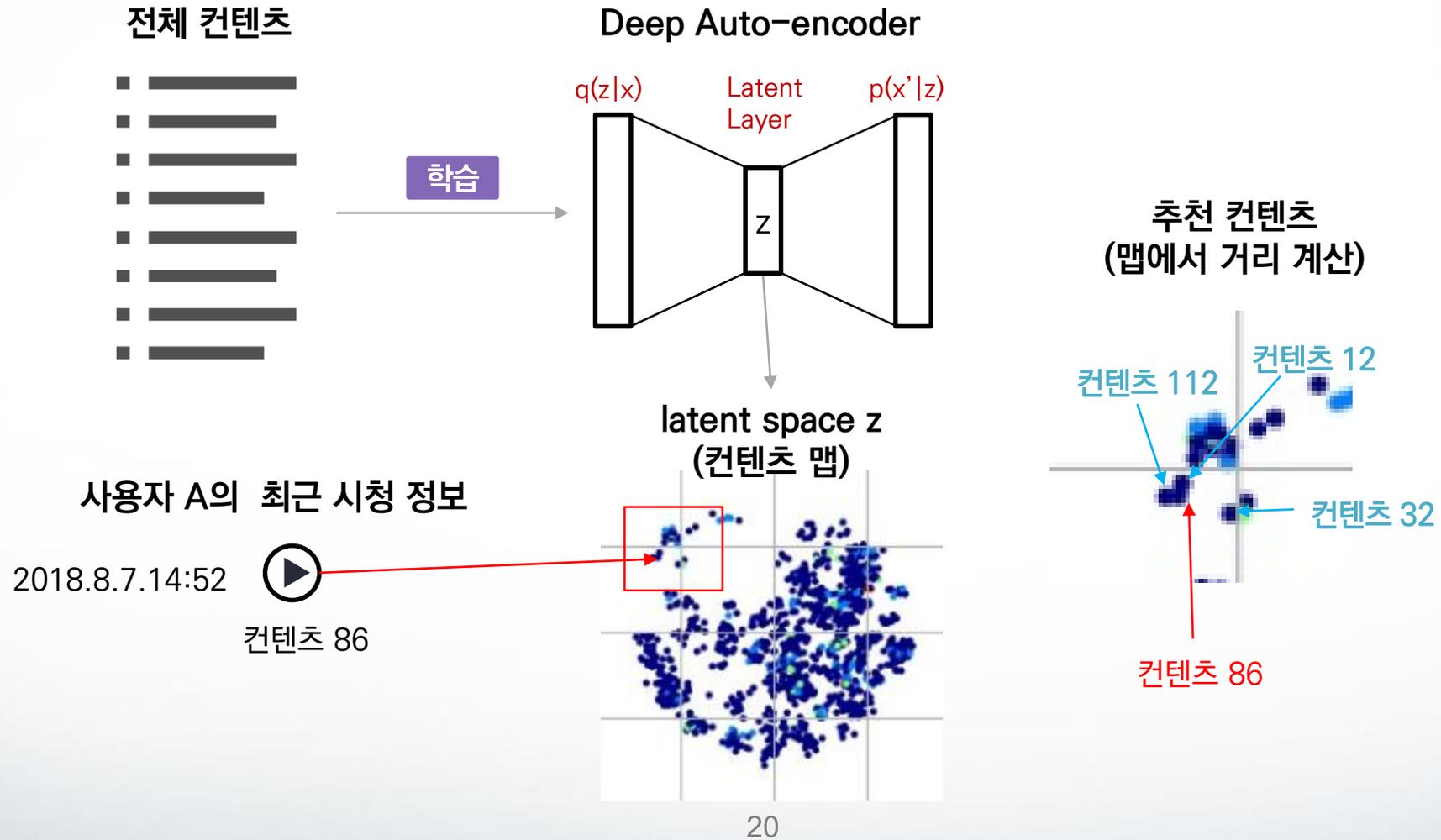
- 비선형 활성화 함수(activation function)를 활용해 사용자간의 관계를 모델링해 기존 선형 학습법에 비해 성능 향상
- 단순 Matrix Factorization에 비해 sparse matrix에서 좋은 성능을 보임
- 평점 추천 (입력이 평점 형태인 경우), 순위 추천 (입력이 1 or 0 인 경우) 에 모두 활용 가능

Deep Auto-encoder로 추천 알고리즘 만들기

실제 추천 사례

1. 콘텐츠 유사도 기반 추천

컨텐츠들의 특성이 반영된 콘텐츠 벡터 공간에서 유사한 컨텐츠들을 추천



무엇을 활용해 콘텐츠를 벡터 공간에 표현할까?

- **콘텐츠 메타정보**

- 콘텐츠 제목
- 콘텐츠 분류
- 콘텐츠 등록일(YYYYMMDD)
- 저자의 이름, 직급, 회사, 부서
- 콘텐츠 길이 (초)
- 콘텐츠 분류

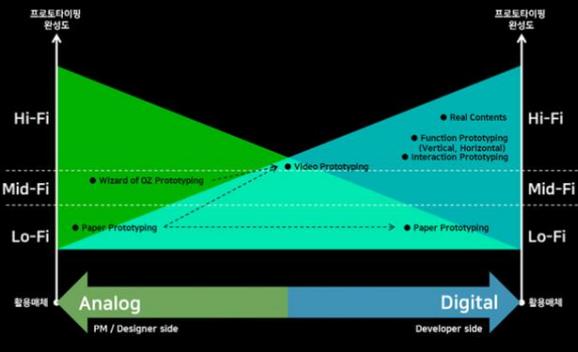
- **콘텐츠 이미지**



프로토타이핑에 대한 학자들의 다양한 정의

뷰라키 홀그림슨 Bjurki Hultgrimsson	새롭고 다양한 신제품에 어떻게 대응할지 연구실험하기 위해 실물을 제작하는 것
존 에드슨 John Edison	제품이나 서비스를 어떤 모습으로 가장 신속하고 간단하게 만들어보는 것
타드 자키 워펠 Tod Zaki Warfel	최종적인 시스템의 시뮬레이션, 혹은 대표 모델

프로토타이핑 분류체계



- **콘텐츠 스크립트**

동영상의 모든 내용이 스크립트로 잘 정리,
 강의 중심의 콘텐츠라 스크립트로 대부분
 내용 파악 가능
 ⇒ 활용 데이터로 선정

컨텐츠 스크립트 전처리

컨텐츠 스크립트를 딥러닝 학습에 적합한 단어-빈도 벡터 형태로 변환

▶ 스크립트에서 단어 추출

```
--TITLE--
디자인 씽킹 - “상상이 실현되는 순간”

--SCRIPT--
지난 시간에는 고객에 대한 이해를 넘어 공감, 소통하며 ‘탁월한 아이디어’를 찾기 위한 방법으로, 프로젝트 상황에 맞는 아이디어 단계별 방법론의 활용에 대해 이야기했습니다. 이번에는 디자인 씽킹 네 번째 단계인 ...
```

명사 추출

```
{
  "title": "디자인 씽킹 상상 실현 순간",
  "script": "지난 시간 고객 이해 공감 소통 탁월 아이디어 방법 프로젝트 상황 아이디어 단계 방법 활용 이야기 이번 디자인 네 씽킹 단계 ..."
}
```

▶ 전체 스크립트에서 유효한 단어 2000개 선정

단어	개수	단어	개수
것	179,481	생각	30,918
수	106,762	시장	24,705
년	66,277	세계	22,870
기업	34,977	중국	21,183
사람	34,719	미국	19,563
...			

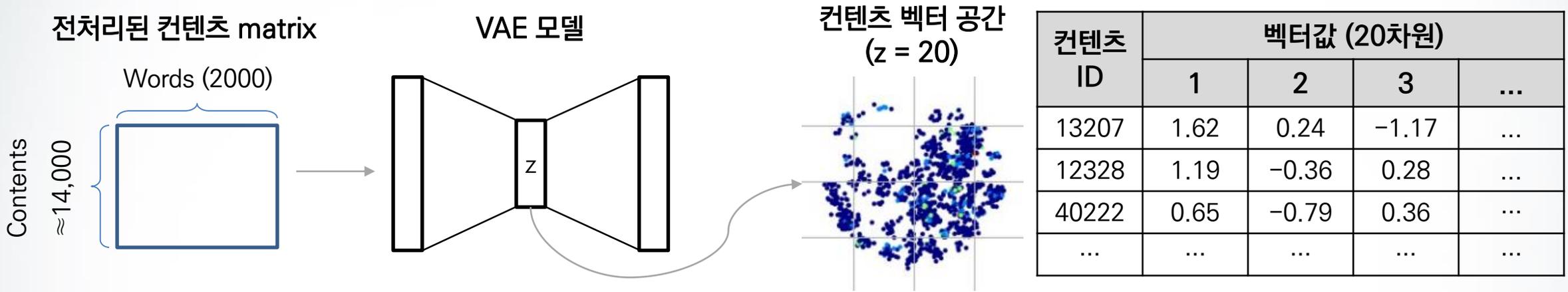
▶ Frequency matrix (Word, Contents)로 제작

2000 단어

		가위	김치	리더	일본	...	사람	미국	기업
컨텐츠1 : 컨텐츠N	컨텐츠1	0.0	0.1	0.0	0.8	...	-0.18	1.24	3.0
	:					...			
	컨텐츠N	0.1	0.0	0.0	1.0	...	1.8	1.2	1.3

딥러닝에 의한 콘텐츠 벡터 공간 학습

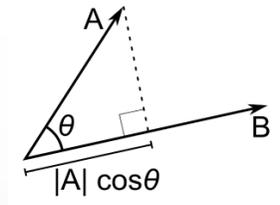
콘텐츠들의 특성이 반영된 콘텐츠 벡터 공간 생성 → 콘텐츠 벡터 공간에서의 근사는 콘텐츠의 유사함을 의미



결과 : 콘텐츠 ID [유사한 콘텐츠 순서]

13207 : [41581 53565 4634 ...]
 12328 : [14786 10036 44143 ...]
 40222 : [66582 13149 14718 ...]
 11233 : [88801 8874 18566 ...]
 ...

콘텐츠 벡터의 거리 측정
(Cosine similarity)



$$\cos(\theta) = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

컨텐츠 유사도 계산에 의한 추천 결과

기준 컨텐츠

물류 거점은 왜, 어떻게 최적화 되어야 하는가?

기준 알고리즘 (ML)

선진물류를 위한 첫걸음, 물류 운영수준 평가 서비스
아는 것이 경쟁력! 화물운송의 안전장치 - 화물배상책임보험, 적하보험
Cello 플랫폼, 인공지능으로 미래를 예측하다
물류 가시성의 숨은 조력자 'Smart Container'

상상이 실현되는 순간
GAN의 활용 사례와 발전방향

물류의 패러다임을 바꾸는 또 하나의 혁신기술, 블록체인!

정적 분석으로 자바스크립트 코드의 오류와 코드 스멜 찾기
소매업 분야에 활용되는 가상현실 : VR/AR이 바꾸는 미래 쇼핑
디자인 씽킹으로 올바른 질문에, 올바른 답을 찾는다.

딤러닝 알고리즘

Cello 플랫폼, 인공지능으로 미래를 예측하다
물류의 패러다임을 바꾸는 또 하나의 혁신기술, 블록체인!
물류 가시성의 숨은 조력자 'Smart Container'
아는 것이 경쟁력! 화물운송의 안전장치 - 화물배상책임보험, 적하보험

선진물류를 위한 첫걸음, 물류 운영수준 평가 서비스
Next물류 - Logistics 4.0은 무엇인가
통합계약을 통한 물류비용 절감

웨어러블 기술을 통한 행사 관리 혁신

신재생 에너지 저장산업과 물류역량

성공적인 IoT 비즈니스 환경 구축의 필요조건

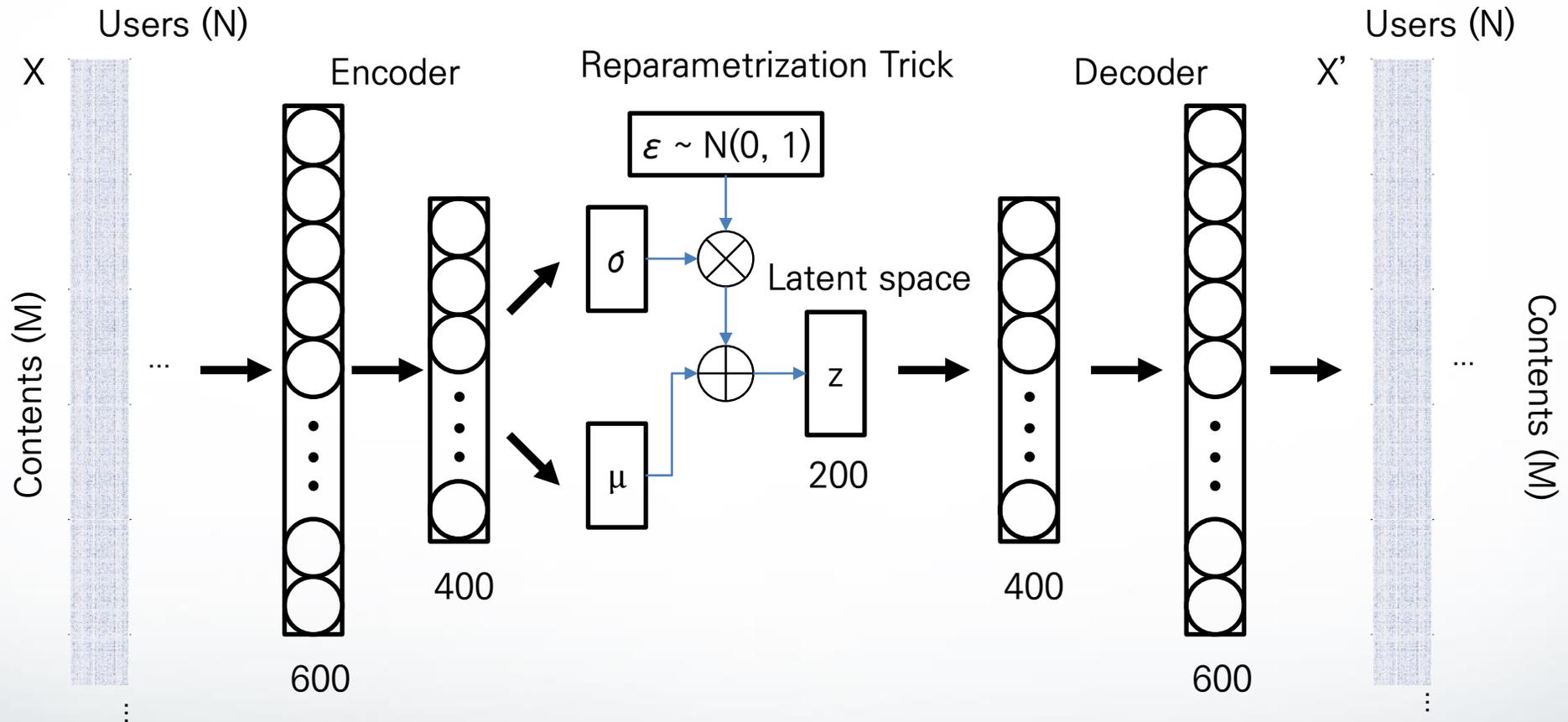
2. 개인화 추천 아이디어 - VAE를 이용한 Matrix completion

학습된 Variational Auto-encoder (VAE)에 사용자-컨텐츠 matrix를 입력했을 때 출력되는 matrix를 통해 컨텐츠를 추천

1. 입력, 학습 데이터 생성을 위한 전처리

2. 모델 생성 및 학습

3. 추천 결과 만들기



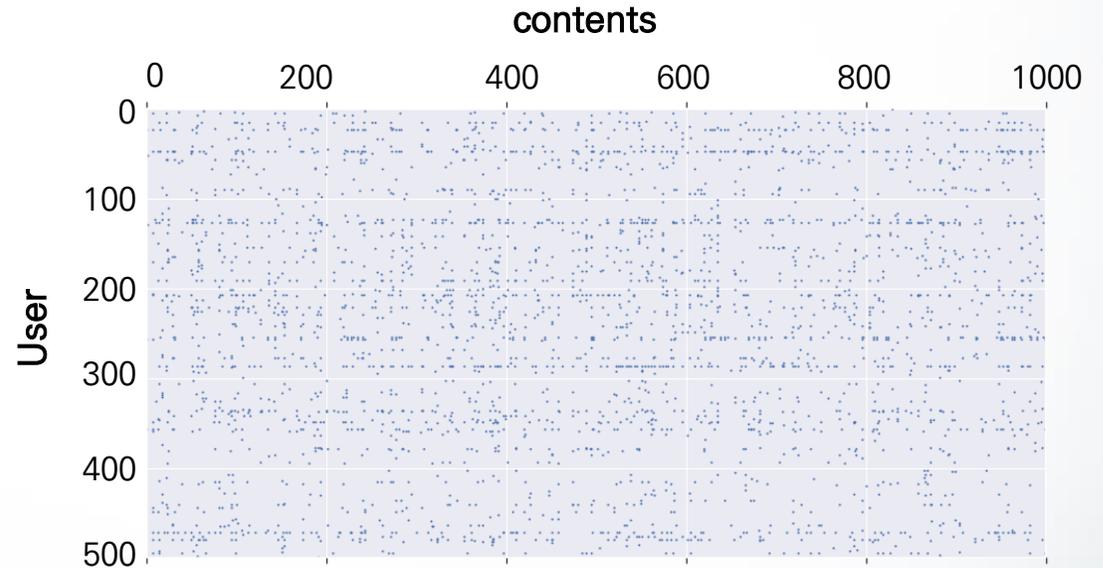
시청 기록 데이터의 전처리

딥러닝이 학습 가능한 형태로 제작

▶ 사용자가 콘텐츠를 시청한 이력

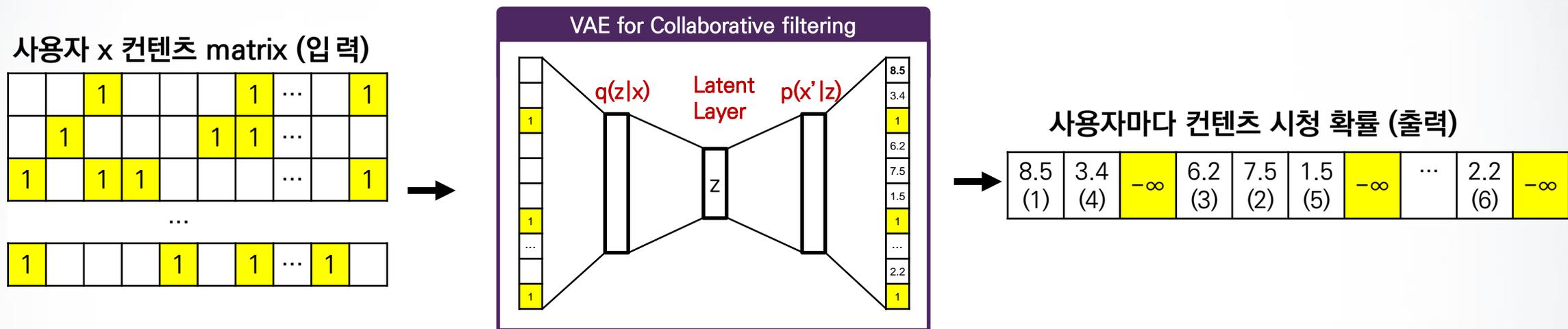
	User Seq.	Cont. Seq.	DateTime	Watch	Cnt	Route	Device	ViewYN
0	10251313	22523	20170517093830	397	1	W	P	Y
1	10251313	77164	20170524140347	454	1	W	P	Y
2	10251313	77244	20170526085524	473	1	W	P	Y
3	10251347	76242	20170515133240	24	1	W	P	Y
4	10251385	76422	20170508065155	3	1	W	P	Y
5	10251385	77742	20170518064041	8	1	W	P	Y
6	10251399	76143	20170525085249	7	1	W	P	Y
7	10251399	77646	20170525085256	4	1	W	P	Y
8	10251399	68085	20170525134501	3	1	W	P	Y
9	10251399	51823	20170530091427	33	1	W	P	Y
10	10251516	74305	20170502163352	3	1	W	P	Y
11	10251516	77548	20170502163438	4	1	W	P	Y
12	10251516	75902	20170502163503	3	1	W	P	Y
13	10251516	67323	20170517182052	4	1	W	P	Y
14	10251516	77282	20170517182102	3	1	W	P	Y
15	10251516	77742	20170517182114	3	1	W	P	Y
16	10251677	76803	20170516080021	640	1	W	P	Y
18	10251677	77242	20170517100248	414	1	W	P	Y

▶ 각각의 시청 이벤트를 토대로 binary matrix 제작 시청 정보가 있으면 1, 없으면 0



사용자별 콘텐츠 추천 결과 출력

학습된 VAE에 사용자-콘텐츠 matrix를 입력했을 때 출력되는 matrix를 통해 콘텐츠를 추천



결과 데이터 출력

사용자 ID	Contents ID / score (상위 1500개)										
12636741	79706/11.787	81382/11.222	82184/10.383	80862/10.190	81350/10.044	75783/9.592	76846/9.366	81122/9.347	81823/9.107	...	17084/3.902
13231441	17630/8.295	88987/8.252	88629/8.231	89627/8.118	88752/8.103	88697/8.092	90751/8.045	88667/8.038	18364/8.018	...	77562/4.561
11992118	36710/7.445	24940/7.223	31066/7.212	28621/7.108	24201/7.065	51444/7.062	25125/7.015	58983/6.942	33162/6.919	...	40504/4.746
13097669	83666/7.728	89227/7.709	86147/7.588	83623/7.524	88751/7.502	89090/7.480	87887/7.461	89667/7.381	87672/7.374	...	66724/4.502
12414486	16343/8.312	16292/8.192	16518/8.132	27988/8.111	16992/8.073	16996/8.044	16689/8.025	16874/7.924	16338/7.864	...	15823/4.009
10707152	79084/5.386	45581/5.364	25078/5.353	44721/5.316	38261/5.292	23344/5.288	33847/5.287	72802/5.285	32569/5.264	...	82986/3.549
12890364	83654/8.595	77722/8.577	76062/8.536	75266/8.488	77022/8.402	77263/8.374	76042/8.277	74142/8.237	71285/8.237	...	22357/5.681
10654443	85973/10.616	86407/10.573	86189/10.070	85907/9.894	85807/9.796	86447/9.769	86587/9.756	85777/9.746	85789/9.676	...	60626/4.840

Deep Auto-encoder로 추천 알고리즘 만들기

Self Q & A

➤ 그래서 어떻게 됐나요?

- 서비스 시작을 위해 노력하는 중!
소수의 관련자 대상 **블라인드 테스트 결과 디퍼닝**이 더 낫다고...

➤ 어려웠던 점은?

- 모델의 성능을 **수치적으로** 나타내기가 어려움
- **90% 이상의 콘텐츠**를 시청한 사용자가 있음 (더 이상 추천 할 것이 없다...)
- 콘텐츠에 **관심 없는 시청자**와,
동영상 재생을 안하고도 **스크립트로 콘텐츠**를 소비한 시청자의 시청 시간은 **모두 0초**

➤ 앞으로 무엇을 해야 할까요?

- 협업필터링 - 초기사용자 문제,
유사 콘텐츠 추천 - 취향이 반영되었다고 보기 어려움
⇒ **하이브리드 모델 개발**이 필요함

감사합니다!

Q & A

Partner

Disrupt

Foresee

