# 패스트 캠퍼스 딥러닝을 활용한 추천시스템 구현 올인원 패키지 Online.

# 파트1. 추천 시스템

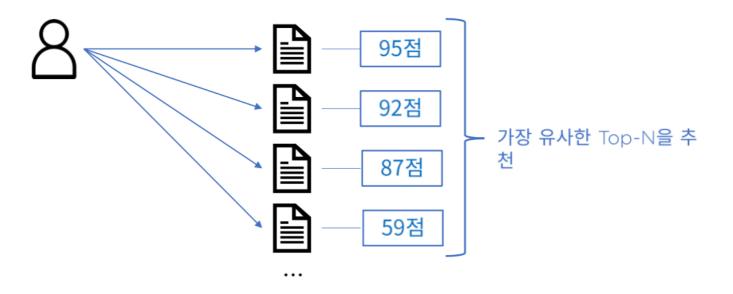
- 사용자와 상품으로 구성된 시스템
- 특정 사용자가 좋아할 상품을 추천
- 비슷한 상품을 좋아할 사용자를 추천
- item + user -> 관심가질만한 정보를 추천

### user and item

- 사용자 : user profile
  - ㅇ 사용자의 고유 정보: 나이, 성별, 지역 등의 개인 신상정보
  - ㅇ 쿠키: 인터넷 방문 기록
  - ㅇ 인터넷 주소
  - ㅇ 웹페이지 방문 기록, 클릭 패턴 등 사용자 행동패턴
  - o 사용자 정보 수집 방법
    - 직접: 설문조사, 평가, 피드백
    - 간점: 웹페이지 머무는 시간, 클릭패턴, 검색 로그 등
    - 개인별 추천 또는 사용자 그룹별 추천 가능
- 아이템 정보 :item profile
  - ㅇ 플랫폼마다 정의하는 아이템의 종류가 다르다.
  - ㅇ 모든 것을 추천하는 플랫폼은 없고 플랫폼마다 관련있는 상품 또는 아이템만 추천한다.
  - ㅇ 아이템 프로필에 속한 정보
    - 아이템 id, 아이템 고유 정보(크기, 색 , 가격 등) , 아이템을 좋아하거나 구매한 사용자 정보 등
- 사용자와 아이템 사이의 관계를 분석하고 연관관계를 찾는다.
- 해당 연관관계를 점수화한다.
- 해당 정보와 아이템 정보를 활용

## 추천 점수란?

- 분석된 사용자와 아이템 정보를 바탕으로 추천 점수를 계산한다.
- 사용자 또는 아이템 프로필에서 어떤 정보를 사용할 지에 따라 추천 알고리즘을 결정한다.
- 사용자 또는 아이템을 추천하기 위한 각각의 아이템 또는 사용자에 대한 정량화된 기준이 필요
- 추천 알고리즘의 목적은 점수화(scoring)하는 것이다.



### 앞으로 배울 내용

주제	소주제
Introduction to Recommender System	추천시스템 소개, 대표 알고리즘, 한계
	추천시스템을 위한 기초 수학, 파이썬 프로그래밍
	추천시스템 평가
	영화 평점과 메타 데이터 분석
Classic Recommender System	컨텐츠기반 추천시스템
	근접이웃기반 협업필터링
	모델기반 협업필터링
Deep Learning based Recommender System	Neural Collaborative Filtering
	Factorization Machine / Deep Factorization Machine
	Wide & Deep Learning for Recommender System
	Autoenoders Meet Collaborative Filtering
	Hybrid & Context-Aware Recommender System
Our Recommender System	텍스트 데이터를 활용한 추천시스템
	이미지 데이터를 활용한 추천시스템
	오디오 데이터를 활용한 추천시스템
	Commercial Recommender System

# 요즘 추천 시스템

추천 시스템의 용도

#### 추천 시스템의 정의

- 사용자 users 로부터 선호도(preference)여부를 데이터화
  - ㅇ 평점 등의 explicit 피드백
  - ㅇ 사용자의 행동 패턴에 대한 implicit 피드백
- 사용자의 선호도와 제한사항을 바탕으로 가장 적절한 아이템 점수를 계산한다
- 사용자에게 가장 관련있는 아이템 순서대로 리스트()

#### 사용

- 정보 검색 분야와 비슷하지만, 비교적 새로운 연구 분야
- 학회
  - o SIGIR, WWW, ACM RecSys, IEEE 등
- 유투브, 넷플릭스, 왓챠, IMDB, 쿠팡, 아마존 , Last.FM, 멜론 등
- Netflix Prize 추천 대회

#### 왜 추천 시스템 이용해야할까?

추천 시스템은 push information, 검색 서비스는 pull information

- 더 많은 아이템 판매
- 더 다양한 아이템 판매: 소비자가 보지 못한 상품을 판매할 수 있다.
- 소비자의 만족도 증가이유, 충성도 높은 고객 등장
- 고객의 니즈 파악
  - ㅇ 데이터가 쌓이면 쌓일 수록 추천 알고리즘의 성능 향상

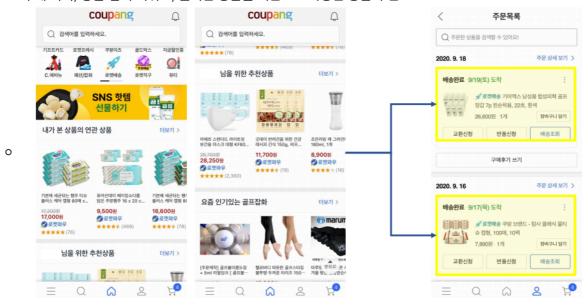
#### 추천 활용 예

- 넷플릭스 : 콘텐츠 추천
  - o 계정 생성시 선택된 좋아하는 컨텐츠 기반으로 콘텐츠 추천
- 인스타그램: 팔로우 추천: 팔로잉하고 있는 계정 유형과 피드에서 자주 클릭한 계정을 분석해 피드에 광고를 노출한다. 계정 팔로우할 겨웅, 해당 계정과 비슷한 계정 추가 추천

•



- 쿠팡: 상품 추천
  - ㅇ 구매 목록, 상품 검색 키워드, 클릭한 상품을 기반으로 다양한 상품 추천



#### 데이터 종류

#### Items

- 추천할 항목
- 가치와 데이터의 복잡함에 따라 구분할 수 있다
- Low: 뉴스, 책, 영화 등
- High: 금융 상품, 직업, 여행 등

#### Users

- 여러 목적과 특징을 갖는 집단
- 행동 패턴 등 으로 유저 모델링
- 협업 필터링: 여러 아이템에 대한 평점 리스트
- 사회적인 요인을 활용한 추천시스 템 등

#### Transactions

- 유저와 아이템간의 상호작용
- 유저가 인터넷에 남긴 로그에서 중요한 정보 추출
- 해시태그, 평점, Implicit한 정보
  모두 활용 가능

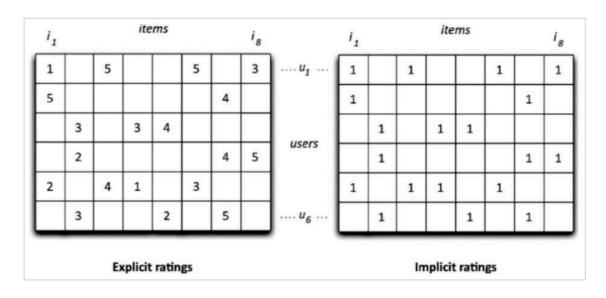
#### 추천 시스템이 풀고자 하는 문제

#### 1. 랭킹 문제

- 1. 특정 유저가 특정 아이템에 대한 평점, 점수를 정확하게 예측할 필요 없음
- 2. 특정 아이템을 좋아할만한 top-k 유저를 선정할 수 있다.
- 3. 특정 유저가 좋아할만한 top-k 아이템을 선정할 수 있다.

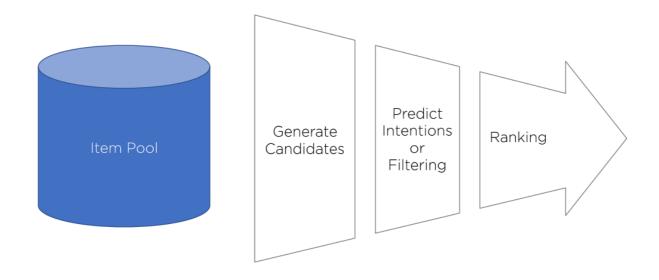
#### 2. 예측 문제

- 1. 유저-아이템 조합에서 평점, 점수 예측
- 2. 유저-아이템 행렬을 채우는 문제
- 3. 유저(m명) 아이템(n개)의 m\*n 행렬, 그러나 비어있는 부분이 존재
  - 1. 관측값(observed value)는 모델 학습에 사용
  - 2. 결측값(missing value)는 모델 평가에 사용



# 추천 시스템 알고리즘

추천 시스템은 랭킹문제 또는 예측 문제이다.

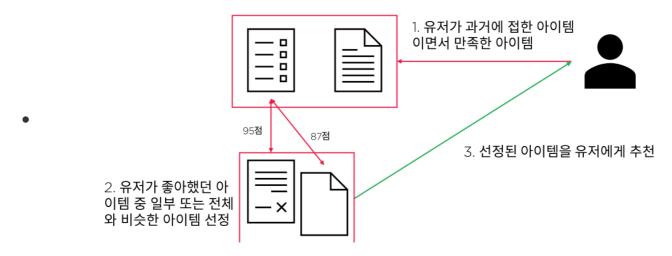


https://towardsdatascience.com/phrase-extraction-in-python-using-lists-tuples-and-lambda-functions-fa0d2eed6dbb

intention: 의도한 바를 추출

## contents-based recommend system: 컨텐츠 기반 추천 시스템

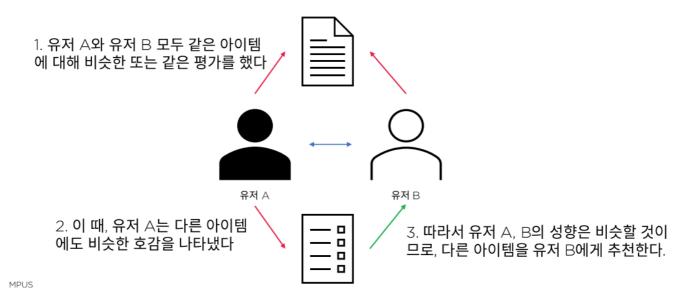
- 사용자가 과거에 좋아했던 아이템을 파악하고, 그 아이템과 비슷한 아이템을 추천한다.
- 예) 스파이더맨 4.5인 유저는 타이타닉보다 캡틴 마블을 더 좋아할거야.



## collaboratice filtering: 협업 필터링

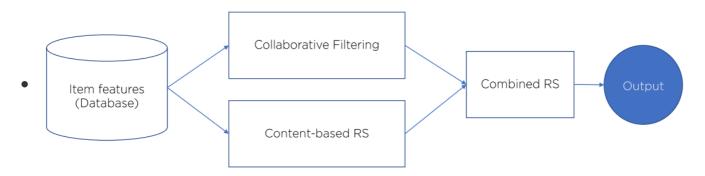
- 비슷한 성향 또는 취향을 갖는 다른 유저가 좋아한 아이템을 현재 유저에게 추천한다.
- 간단하고, 수준 높은 정확도

예시) 스파이더맨에 4.5점을 준 2명의 유저 → 유저 △가 과거 좋아했던 캡틴 마블을 유저 B에게 추천



## hybrid recommend system

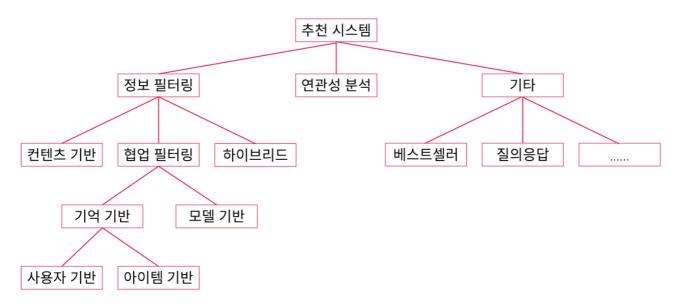
- content based 와 collaborative filtering 의 장단점을 상호보완
- collaborateive filtering 은 새로운 아이템에 대한 추천 부족
- 이에, content-based 기법이 cold-start 문제에 도움을 줄 수 있다.

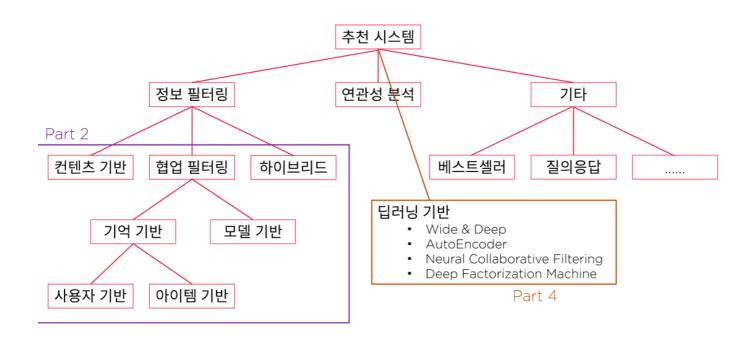


# other recommendation algorithms

- 1. context-based recommendation
  - 1. context-aware recommendation system : 맥락 기반
  - 2. Location-based recommendation system
  - 3. Real-time or time-sensitive recommendation system
- 2. Community-based recommendation

- 1. 사용자의 친구 또는 속한 커뮤니티의 선호도를 바탕으로 추천
- 2. sns 등의 뉴스피드 또는 sns 네트워크 데이터 등 활용
- 3. Knowledge-based recommendation
  - 1. 특정 도메인 지식을 바탕으로 아이템의 features를 활용한 추천
  - 2. Case-based recommendation
    - 1. 사용자의 니즈(현재 문제 등)와 해결책 중 가장 적합한 것을 골라서 추천
  - 3. Constraint-based recommendation
    - 1. 사용자에게 추천할 때, 정해진 규칙을 바탕으로 추천

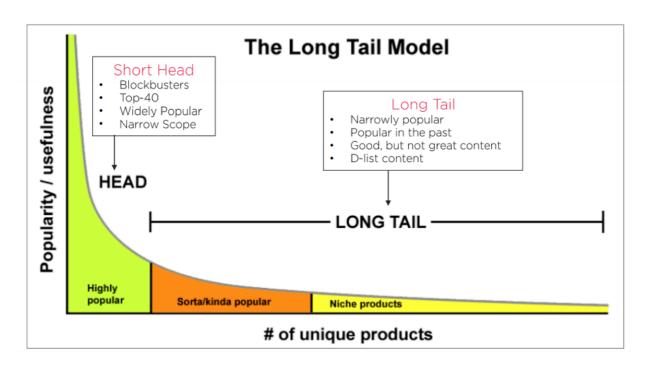




# 추천 시스템의 한계

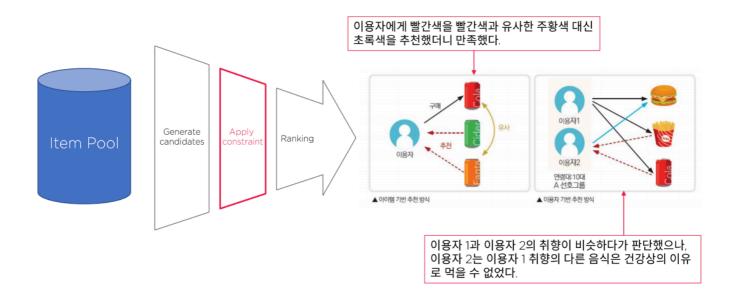
- scalabilty
  - ㅇ 실제 서비스 상황은 다양한 종류의 데이터
  - ㅇ 학습 또는 분석에 사용한 데이터와는 전혀 다른 실전 데이터( 코로나, 바뀐 유행)
- proactive recommend system
  - o 특별한 요청 없이도 사전에 먼저 제공하는 추천 서비스(push service)
  - ㅇ 모바일, 인터넷 등 어디서는 유저에게 끊임없이 신선한, 좋은 정보를 추천할 수 있는 서비스
- cold-start problem
  - ㅇ 추천서비스를 위한 데이터 부족
  - ㅇ 기본적인 성능을 보장하는 협립필터링 모델 구축이 쉽지 않은 것이 일반적
  - ㅇ 컨텐츠 기반 또는 지식 기반의 방법 역시 서비스로 적용하기 어려움.
- privacy preserving recommender system
  - ㅇ 개인 정보 등 유저 정보가 가장 중요하지만, 직접 사용하기 어려움.
- mobile devices and usage context
  - personal computing, location based services(LBS)
  - ㅇ 개별상황 또는 환경 등에 따라 다른 컨텍스트를 사용
- long-term and short term user preference
  - o 개인 또는 그룹의 장기.단기 관심사항(예전은 맞지만,지금은 아니다.)
  - ㅇ 추천받고 싶은 아이템이 현재 또는 과거 중 어느 시기와 관련있는지 파악하기 어렵다.
- generic user models and cross domain recommender system
  - ㅇ 하나의 모델을 여러가지 데이터에 적용하기 어렵다.
  - ㅇ 비슷한 도메인의 데이터를 활용해도 동일한 성능의 추천 시스템을 기대하기 어렵다.
- starvation and diversity
  - o starvation: 필요한 컴퓨터 자원을 끊임 없이 가져오지 못하는 상황
  - ㅇ 유저, 아이템이 다양하고, 모든 유저, 아이템에 더 많은 관심을 부여해야한다.

### long tail 경제



short head 는 누구나 유명하고, 적지만 (10) 알려지지 않은 long tail 이 있다. (90) 양쪽을 고려하는 고민을 해야한다.

### 추천 시스템엔 답이 없다



# numpy & pandas