

# 알고리즘 마케팅 10강

2023. 4. 27. (목)

서울과학기술대학교 데이터사이언스학과

김 종 대

# 오늘의 강의

- 9주차 Review
- 자연어 처리: 워드 임베딩
- 추천 시스템 개관
- 콘텐츠 기반 필터링
- 협업 필터링
- 마케팅 모형 기반의 추천 시스템

# 검색

- **검색**

- 목적: 검색 창이나 선택된 필터로 표현되는 고객의 검색 의도에 따른 결과물을 제시
- 개인화된 솔루션을 제공한다는 점에서 추천 서비스와도 유사한 점이 많음.

- **검색의 비즈니스 목표**

- 일반적 목표: 사용자의 의도를 이해하고 그 의도에 적합한 결과를 전달하는 것
- 1. 적합성: 고객의 의도와 일치하는 검색 결과를 얼마나 정확하게 제시하는가?
- 2. 상품 통제: 마진에 따른 순위 조정 또는 고객 프로파일에 따른 필터링
- 3. 검색 성능 지표: 컨버전 비율, 검색어 수정 비율, 페이지 비율, 검색 대기 시간 등

# Bayesian Statistics

- **Why Bayesian statistics?**
  - To express our information and beliefs about unknown quantities
  - Bayes' rule provides a rational method for updating beliefs in light of new information.
- Bayesian methods provide:
  - Parameter estimates with good statistical properties (e.g. distribution information of parameter set)
  - Parsimonious descriptions of observed data
  - Predictions for missing data and forecasts of future data
  - A computational framework for model estimation, selection and validation

# Bayesian Statistics

- Bayes' rule

$$\Pr(\theta|y) = \frac{\Pr(y|\theta) \Pr(\theta)}{\Pr(y)}$$

- $\Pr(y|\theta)$ : 특정 parameter set이 주어졌을 때 데이터  $y$ 가 생성될 확률을 반영한 모형 (sampling model)
- $\Pr(\theta)$ : 특정 parameter set에 대한 기존에 누적된 주관적 믿음, 즉 parameter에 대한 사전 확률(prior distribution)
- $\Pr(\theta|y)$ : 기존에 누적된 주관적 믿음에 새로운 데이터  $y$ 를 업데이트한 결과, 즉 베이저안 통계학에서 최종적으로 추정하고자 하는 사후 확률(posterior distribution)
- $\Pr(y) = \int \Pr(y|\tilde{\theta}) \Pr(\tilde{\theta}) d\tilde{\theta}$ : 일종의 상수로서 일반적으로 numerical method를 통해 계산하여야 함.

# Bayesian Statistics

- **Summary: Numerical methods for Bayesian Statistics**
  - Conjugate prior  $\Rightarrow$  We can easily get the analytic solution (Known form of probability distribution).
  - Semiconjugate prior  $\Rightarrow$  **Gibbs sampler** (using the closed form of full conditional distribution)
  - Nonconjugate prior  $\Rightarrow$  **Metropolis-Hastings algorithm**
- In order to estimate latent Dirichlet allocation model, we usually use **Gibbs sampler** methods.

# Bayesian Statistics

- **Hierarchical model using a Bayesian approach**
  - We can model additional layers by probability distributions for parameters.
  - Example: normal distribution with  $(\mu, \sigma^2)$ 
    - Assume that we need to propose a model to figure out differences between groups (j).
    - For given  $\sigma^2$ , we can model  $\mu$  as  $\Pr(\mu_j | \varphi) = \text{Normal}(\theta, \tau^2)$  where  $\varphi = \{\theta, \tau^2\}$ .
    - We can add another layers for  $\varphi = \{\theta, \tau^2\}$ .

# Bayesian Statistics

- **Gibbs sampler**
  - Gibbs sampling uses distribution information of full conditional distribution which is a conditional distribution of a parameter given everything else.
  - For normal distribution  $x \sim N(\theta, \sigma^2)$ , full conditional distributions are:
    - $\Pr(\theta | \sigma^2, y_1, \dots, y_n)$  for mean parameter
    - $\Pr(\sigma^2 | \theta, y_1, \dots, y_n)$  for variance parameter
  - By using these full conditional distributions, we can make the iterative sampling idea.



# Bayesian Statistics

- **Gibbs sampler**

- Given a current (or initial) state of parameters  $\phi^{(s)} = \{\theta^{(s)}, (1/\sigma)^{2(s)}\}$ , we generate a new state as follows:
  - 1. Sample  $\theta^{(s+1)} \sim \text{Pr}(\theta | (1/\sigma)^{2(s)}, y_1, \dots, y_n)$
  - 2. Sample  $(1/\sigma)^{2(s+1)} \sim \text{Pr}((1/\sigma)^2 | \theta^{(s+1)}, y_1, \dots, y_n)$
  - 3. Let  $\phi^{(s+1)} = \{\theta^{(s+1)}, (1/\sigma)^{2(s+1)}\}$
- Likewise, the Gibbs sampler generates a dependent sequence of parameters  $\{\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(S)}\}$ .
- S개의 parameter sample들의 평균 등으로 parameter의 추정치를 계산
  - $E[\theta | y_1, \dots, y_n] \approx \frac{1}{S} \sum_{s=1}^S \theta^{(s)}$

# Bayesian Marketing

- Allenby, G. M., & Rossi, P. E. (1998). Marketing models of consumer heterogeneity. *Journal of econometrics*, 89(1-2), 57-78.
  - The distribution of consumer preferences plays a central role in many marketing activities.
  - Marketing activities which target specific households (individuals) require household (individual) level parameter estimates.
  - Thus, the modeling of consumer heterogeneity is the central focus of many statistical marketing applications.

# Bayesian Marketing

- Allenby, G. M., & Rossi, P. E. (1998). Marketing models of consumer heterogeneity. *Journal of econometrics*, 89(1-2), 57-78.
  - A Bayesian approach to estimate  $\alpha_i$  (individual level parameter) and  $\beta$  (parameters for the distribution of  $\alpha_i$ ) by maximizing the following equation

$$\Pr(\alpha_i | data) \propto \Pr(data_i | \alpha_i) \pi(\alpha_i | \beta = \hat{\beta})$$

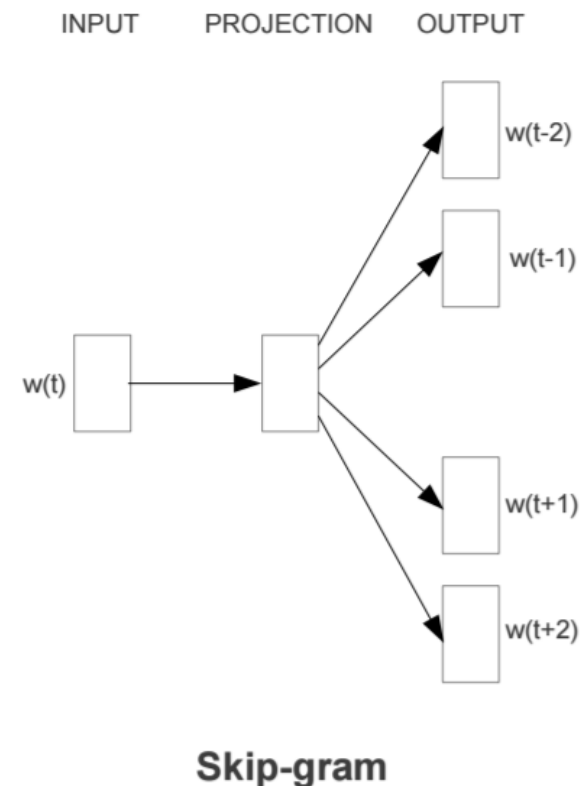
- Incorporating heterogeneity in choice models

$$\beta_i \sim iid N(\bar{\beta}, V_{\beta})$$

$$\bar{\beta} \sim N(\bar{\beta}, aV_{\beta}), \quad V_{\beta}^{-1} \sim W(v_0, V_0)$$

# Key Models for Text Mining

- 텍스트 마이닝 모형에 대한 여러 분류 방식이 있겠으나, 여기에서는 단어 표현 방법에 따른 분류로 소개
- 단어의 표현 방법 (Word representation)
  - 국소 표현 (Local representation): 해당 단어 자체로 값을 할당
    - 주로 사전에 정의한 의미 사전(dictionary)을 기반으로 값을 할당
  - 분산 표현 (Distributed representation): 단어를 표현할 때 함께 사용된 주변의 단어까지 참고하는 방식
    - 단어의 맥락적 의미나 뉘앙스를 비교적 잘 반영할 수 있다는 장점



# Local Representation-Based Approach

- **Bag of Words**

- 문서 내에 들어있는 단어를 하나의 가방(bag)에 전부 집어넣는 방식
- 단어의 출현 빈도(frequency)에 주목하는 방식으로, 단어의 순서나 맥락을 반영하지 못함.
- 한계: 단어의 맥락적 의미가 중요하거나, 한 문서 내에 여러 주제나 의견이 들어있을 경우 문서 내의 유의미한 의미를 제대로 포착하지 못할 수 있음.

- **Document-Term Matrix (DTM)**

- Bag of words는 단어의 빈도 수 기반이므로 각 문서(행)와 각 단어(열)로 이루어진 행렬로 표현

# TF-IDF

- **TF-IDF (Term frequency-inverse document frequency)**
  - 단어의 빈도 + 역 문서 빈도를 이용하여 문서 내 특정 단어의 중요도를 구하는 작업 등에 사용
  - 문서 집합 내 **핵심어 추출**이나 **문서 간 유사도**를 계산하는 데 응용 가능
  - 문서 내에 등장하는 모든 단어에 대해 빈도를 표시하는 것이 비효율적이라면, 중요한 단어에 대해 가중치를 줄 수 있음.
- 역 문서 빈도(Inverse document frequency): 특정 단어  $w$ 가 등장한 문서의 수에 반비례하는 값
- 원리: 불용어(the, it, as 등)는 거의 모든 문서에 자주 등장하게 됨. 즉, 일부 문서에만 등장하는 일종의 “희귀 단어”들이 의미 파악에 중요하고, 거의 모든 문서에 등장하는 단어들은 중요성이 떨어질 것이라고 가정하여 단어의 중요도 가중치를 조정하는 방법

# Latent Dirichlet Allocation (LDA)

- **토픽 모형 (Topic Model)**
  - **잠재 디리클레 할당 모형 (Latent Dirichlet Allocation; LDA)**
    - 장점 1. 베이저안 확률 모형을 활용하기 때문에 확률 정보를 이용 가능
    - 장점 2. 문서의 집합을 구성하는 토픽의 수와 토픽별 의미를 추론할 수 있음.
    - 장점 3. 각 문서가 어떤 토픽으로 어떻게 구성되어 있는지 확률적으로 표현할 수 있음.
  - 단점 1. Bag of words 기반이기 때문에 단어의 맥락적 의미를 충분히 반영할 수 없음.
  - 단점 2. 토픽별 의미는 키워드를 바탕으로 연구자 등이 직접 해석해야 함.

# Latent Dirichlet Allocation (LDA)

- **Generative process**
  - For document  $w$  in a corpus  $D$ ,
    - 1. Choose  $N$  (the number of words)  $\sim \text{Poisson}(\xi)$  (or choose  $N$  from data)
    - 2. Choose  $\theta \sim \text{Dirichlet}(\alpha)$
    - 3. For each of the  $N$  words  $w_n$ :
      - (a) Choose a topic  $z_n \sim \text{Multinomial}(\theta)$
      - (b) Choose a word  $w_n$  from  $\text{Pr}(w_n|z_n, \beta)$ , a multinomial probability conditioned on the topic  $z_n$ .



# Distributed Representation-Based Approach

- **BERTopic Model (Grootendorst, 2022)**
  - Sentence-level BERT 기반의 토픽 모형 (문서 클러스터링 모형)
    - Bidirectional Encoder Representations from Transformers (BERT)
  - 1. SBERT: 주어진 텍스트 데이터에 대해 문서 단위의 임베딩 진행
  - 2. HDBSCAN (Hierarchical density-based spatial clustering of applications with noise): 유사한 문서 간의 클러스터링 진행 (하나의 클러스터 = 하나의 잠재 토픽)
  - 3. UMAP (Uniform manifold approximation and projection): 데이터 차원 축소에 활용
  - 4. c-TF-IDF (A class-based version of TF-IDF): 토픽별 키워드 추출 및 제시

# Text Preprocessing

- 1. 토큰화 (tokenization)
  - 언어학의 음운론, 형태론에 해당하는 단계로서 텍스트를 단어나 형태소 등의 단위로 쪼개는 작업
- 2. 정제 및 정규화 (cleaning and normalization)
  - 표기법 통일(예: USA, United States), 대소문자 통일
- 3. 불용어 처리 (stopword)
  - 관사, 전치사 등 텍스트 의미 분석에 큰 의미가 없는 단어를 제거
- 4. 어간 (stemming) 및 표제어 (lemmatization) 추출
  - 단어의 의미를 담고 있는 핵심 부분을 추출
- 참고: 정규표현식을 이용한 전처리 (regular expression)

# Regular Expression

- 정규표현식을 이용한 전처리 (regular expression)
  - 문자의 공통된 패턴을 이용해 텍스트 데이터를 정제할 정규적인 규칙을 부여

규칙	설명
\\	backslash (\)
\d	모든 숫자(digits)
\D	숫자를 제외한 모든 문자, [^0-9]와 동일한 의미
\s	모든 공백(space), [\t\n\r\f\v]와 동일한 의미
\S	공백을 제외한 모든 문자, [^\t\n\r\f\v]와 동일한 의미
\w	문자 또는 숫자, [a-zA-Z0-9_]와 동일한 의미
\W	문자 또는 숫자가 아닌 문자, [^a-zA-Z0-9_]와 동일한 의미

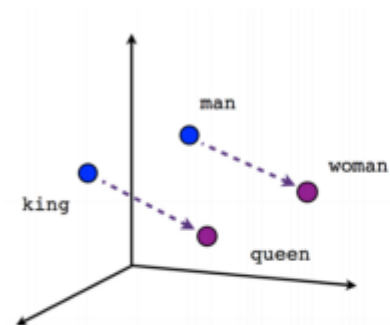
# Local Representation-Based Approach

- **Joint Sentiment Topic Model (JST)**

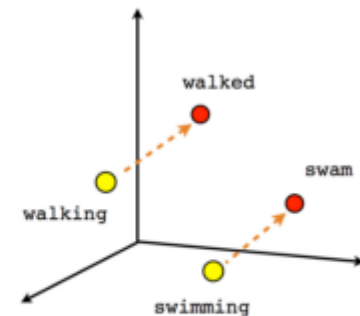
- 감성 분석과 토픽 모형 분석은 마케팅, 비즈니스 애널리틱스 분야에서 매우 유용하게 사용
- 그러한 맥락에서 두 모형을 동시에 적용할 수 있는지에 대한 방법론적 확장이 관심사
- Two-step method: 토픽 모형의 추정 결과로 나온 키워드들을 기반으로 추가적인 감성 분석을 시행
  - 특정 토픽이 긍정적인지 부정적인지 판별하는 것이 가능
- One-step method: 토픽 모델링의 관점에서, 기존의 베이지안 모형 구조에 감성(sentiment)에 대한 층위를 추가하여 방법론적으로 확장한 것이 바로 JST 모형!

# Distributed Representation-Based Approach

- 분산 표현(distribution representation) 기반의 모형:  
대다수가 머신 러닝 또는 딥러닝 기반의 모형
- 워드 임베딩 (Word embedding)
  - 각 단어를 하나의 벡터(예: [300 X 1] 벡터)로 할당
  - 텍스트 데이터를 이용해 각 단어의 맥락적 의미를 반영하는 의미 벡터 공간을 추정하고, 각 단어의 의미를 의미 벡터 공간에서의 기하학적 위치로 표현
  - 예: LSA, Word2vec, FastText, Glove 등



Male-Female



Verb tense

# Distributed Representation-Based Approach

- **Word2vec** 모형

- 워드 임베딩을 구현하는 대표적인 방법론으로서, 언어학의 분포 가설에 기반을 둠.
- 분포 가설(distributional hypothesis): 비슷한 맥락에 등장하는 단어는 비슷한 의미를 가진다.
  - ⇒ 텍스트 데이터 내에서 더 많은 맥락(문장에서의 위치 등)을 공유하는 단어일수록 의미가 더 비슷할 것이라고 간주
- 분포 가설을 바탕으로 의미가 비슷하다고 판단되는 단어들을 다차원 벡터공간 상에 가깝게 위치하도록 할당
- **Sentence2vec**: 단어가 아닌 각 문장을 벡터공간에 할당하는 모형
- **Doc2vec**: 단어가 아닌 각 문서를 벡터공간에 할당하는 모형

# Word Embedding

- Word Representation

- Sparse Representation

- One-Hot Encoding: 단어 전체 집합의 크기(N)를 차원으로 하는 벡터로 단어를 표현하되, 해당 단어의 인덱스만 1을 부여하고 나머지는 0을 부여하는 표현 방식
    - 예: {점심: [0, 1, 0, 0, 0, ... , 0, 0, 0], 저녁: [0, 0, 0, 1, 0, ... , 0, 0, ], ...}
    - 예: Document-Term Matrix
  - 한계 1. 저장 공간 관리 면에서 매우 비효율적 (N이 매우 크다면?)
  - 한계 2. 단어의 유사도를 표현하는 것이 불가능

# Word Embedding

- **Word Representation**

- Dense Representation

- Sparse representation과 달리, 단어 전체 집합의 크기(N)가 아닌 연구자가 사전에 부여한 값으로 차원의 크기를 설정
    - 데이터 학습을 통해 0, 1 외의 실수 값을 부여하여 좀 더 밀집된 형태로 단어를 벡터화
    - 예: {점심: [0.3, 2.0, -1.7, ... , 5.7, 2.8], ...}
  - Word embedding: 단어를 dense representation으로 나타내는 방법론
  - LSA, Word2Vec, FastText 등이 word embedding에 해당한다고 볼 수 있음.



# Word2Vec

- Word2Vec

- Distributed Representation

- 분포 가설 (distributional hypothesis): *비슷한 문맥에서 등장하는 단어들은 비슷한 의미를 가진다.*
    - 텍스트 데이터를 이용해 각 단어의 맥락적 의미를 반영하는 의미 벡터 공간을 추정하고, 각 단어의 의미를 의미 벡터 공간에서의 기하학적 위치로 표현
    - 즉, 텍스트 데이터에서 더 많은 맥락을 공유하는 단어일 수록 더 비슷한 의미를 가질 것으로 간주하고 의미 벡터 공간에 가까운 위치에 할당

- Algorithms for Word2Vec

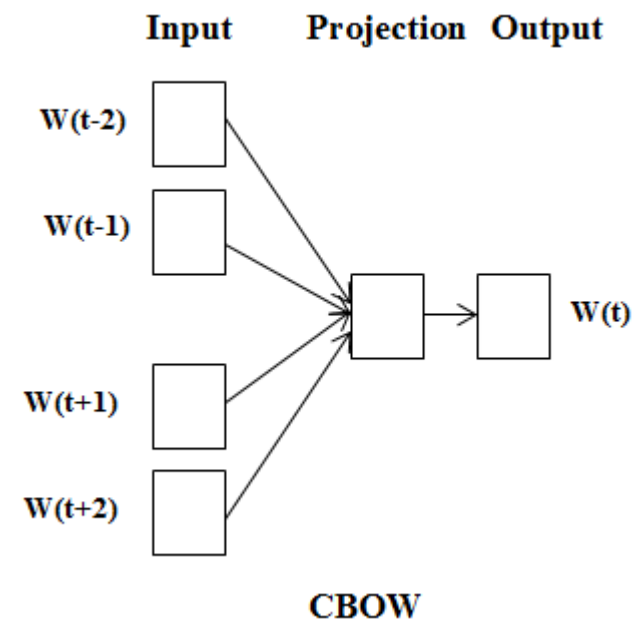
- Continuous Bag of Words (CBOW)
    - Skip-Gram

# Word2Vec

- Word2Vec

- Continuous Bag of Words (CBOW)

- 주변 단어(context word)를 이용해 중심 단어(center word)를 학습, 예측
    - 주변 단어를 몇 개 고려할 것인가, 즉 학습시킬 맥락의 사이즈를 얼마나 할 것인지 윈도우(window) 값을 부여
    - Sliding window: 윈도우를 이동하면서, 즉 중심 단어와 주변 단어 조합을 바꿔가면서 데이터 학습을 진행

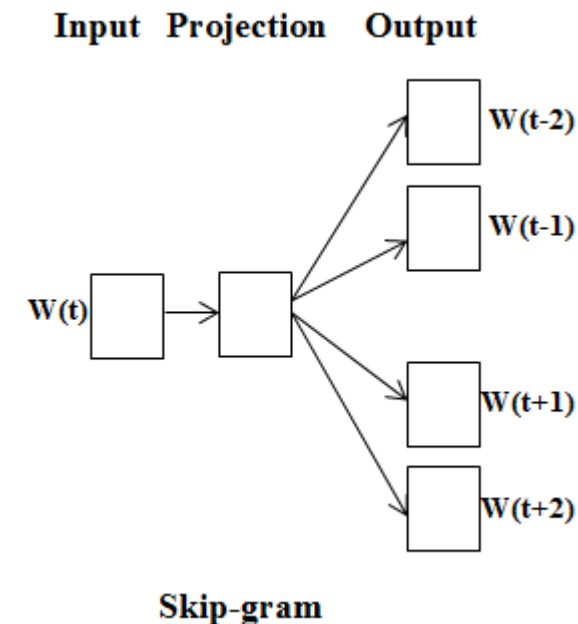


# Word2Vec

- Word2Vec

- Skip-Gram

- 중심 단어(center word)를 이용해 주변 단어(context word)를 학습, 예측
    - 주변 단어를 몇 개 고려할 것인가, 즉 학습시킬 맥락의 사이즈를 얼마나 할 것인지 윈도우(window) 값을 부여
    - Sliding window: 윈도우를 이동하면서, 즉 중심 단어와 주변 단어 조합을 바꿔가면서 데이터 학습을 진행
    - 일반적으로 CBOW보다 우수한 성과를 보고한다고 알려져 있음.



# 추천 (Recommendation)

## ■ 추천 시스템

- 많은 대중이 선호하는 인기 제품 외에 니치 제품에 대한 수요도 분명히 존재
- 롱테일(*long tail*): “니치 제품에 대한 수요의 합은 인기 제품의 전체 수요만큼이나 커질 수 있다.”
- 온라인 채널의 등장으로 이러한 니치 제품에 대한 적극적인 마케팅도 가능한 시대 도래
  - 수많은 니치 제품이 제안될 수 있으나, 실제 고객이 찾아보는 제품은 전체 제품 중 극히 일부
  - 제품의 발견, 즉 추천 서비스의 필요성 대두

## The Long Tail



# 추천 (Recommendation)

- 추천 시스템

- 추천 서비스의 기본적인 세팅은 검색 서비스와 유사
  - 고객에게 아이템의 정렬된 리스트를 제공하는 것이 목표
- 고객 평점(rating)이 추천에 있어 가장 핵심적인 데이터
  - 이때 평점은 고객이 직접 입력할 수도 있고, 다른 고객 데이터를 이용해 추정할 수도 있음.
  - 간접적인 평가 결과로서 아이템의 유사성이나 고객의 유사성이 함께 활용될 수 있음.
  - 추천 요구와 고객 평점은 시간, 장소, 마케팅 채널 등과 같은 맥락 정보와도 결합 가능함.

# 추천 (Recommendation)

- 추천 시스템의 비즈니스 목표

- 추천 서비스의 비즈니스 목표는 기본적으로 검색 서비스의 목표와 유사
- 중요한 차이는 추천 시스템에서 고객의 검색 의도는 명백히 표현되지 않고 아예 존재하지 않을 수도 있다는 것
  - 즉, 검색은 ‘적합한’ 결과를 제공해야 한다는 목표가 있으나 추천은 ‘적합성’의 기준이 애초에 명시적으로 존재하지 않을 수 있음.
- 비즈니스 목표: *참신함, 유연성, 다양성*

# 추천 (Recommendation)

- 추천 시스템의 비즈니스 목표

- 1. 참신함

- 사용자에게 알려지지 않았던 옵션을 제공
    - 사용자가 이미 알고 있을 가능성이 높은 인기 있는 아이템의 추천은 무의미할 수 있음.

- 2. 우연성

- 고객이 기대하지 않았던 놀라운 제품을 발견하게끔 도와줄 수 있음.
    - 과학적 모형을 통한 예측을 통해 추천을 하면(예: 아예 다른 카테고리의 추천), 고객 입장에서 이것이 ‘놀라운 우연’으로 인식될 수 있을 것
    - 추천 서비스와 고객 간의 장기적인 관계를 구축하는 데 기여

# 추천 (Recommendation)

- 추천 시스템의 비즈니스 목표

- 3. 다양성

- 추천 리스트는 구매 기회를 증가시킬 수 있게 다양해야 함.
    - 지나치게 유사성이 높은 아이템들로만 이뤄진 추천 리스트는 적합도가 높고 유연성이 높게 지각된다고 하더라도 플랫폼 입장에서는 최적이지 아닐 가능성이 있음.



# 추천 (Recommendation)

- 추천 품질 평가

- 추천 서비스의 품질을 평가할 수 있는 정량적 지표에는 무엇이 있는가?
- 검색 서비스의 경우, 전문가에 의한 판단이 가능하지만, 추천 서비스는 고객 프로파일마다 다른 결과가 도출되므로 전문가에 의한 일률적인 평가가 어려움.
- 추천 문제를 고객 평점(rating)에 대한 예측 문제로 간주하고 이를 정량적으로 평가 가능
- 1. 예측 정확도
  - 일반적인 예측 모형 평가 지표: MSE, RMSE, NRMSE

# 추천 (Recommendation)

- 추천 품질 평가

- 2. 랭킹 정확도

- 추천 리스트 상위에 정렬된 K개의 아이템에 대한 평가
    - 정밀도 (precision): K개 중 적합한 아이템의 비율
    - 재현율 (recall): K개 중 실제 사용한(혹은 클릭한) 아이템의 비율
  - 추천 리스트가 짧으면 적합한 아이템을 놓쳐 정밀도가 낮아질 위험 존재
  - 반대로 추천 리스트가 너무 길면 부적합한 아이템을 많이 포함시킬 위험 존재

# 추천 (Recommendation)

- 추천 품질 평가

- 3. 참신성

- “참신하다”의 기준: 추천이 제공될 시점에 고객이 추천된 아이템을 알지 못했을 경우
    - 이는 관측되는 데이터가 아니므로 설문조사를 하거나 모형을 통해 예측해야 하는 문제
      - 모형 가정의 예: 훈련 데이터의 시간 경계( $t$ ) 직후에 랭크/구매된 아이템을 예측하는 것은 보다 먼 미래에 랭크/구매된 아이템을 예측하는 것보다 낮은 참신함을 제공 → 짧은 시간 내에 바로 구매하는 아이템은 고객에게 이미 알려져 있던 아이템일 가능성이 높다는 것

# 추천 (Recommendation)

- 추천 품질 평가

- 4. 우연성

- “우연하다”의 기준: 다분히 주관적이어서 “참신하다”보다 정량화가 어려운 문제
    - 특정한 가정에 기반을 둔 휴리스틱을 활용 가능
      - 예: 우연성을 모형화한 평가용 알고리즘에 의해 개발된 추천과 우연성이 없는 아이템으로 이루어진 기본적인 알고리즘의 추천을 비교

# 추천 (Recommendation)

- 추천 품질 평가

- 5. 다양성

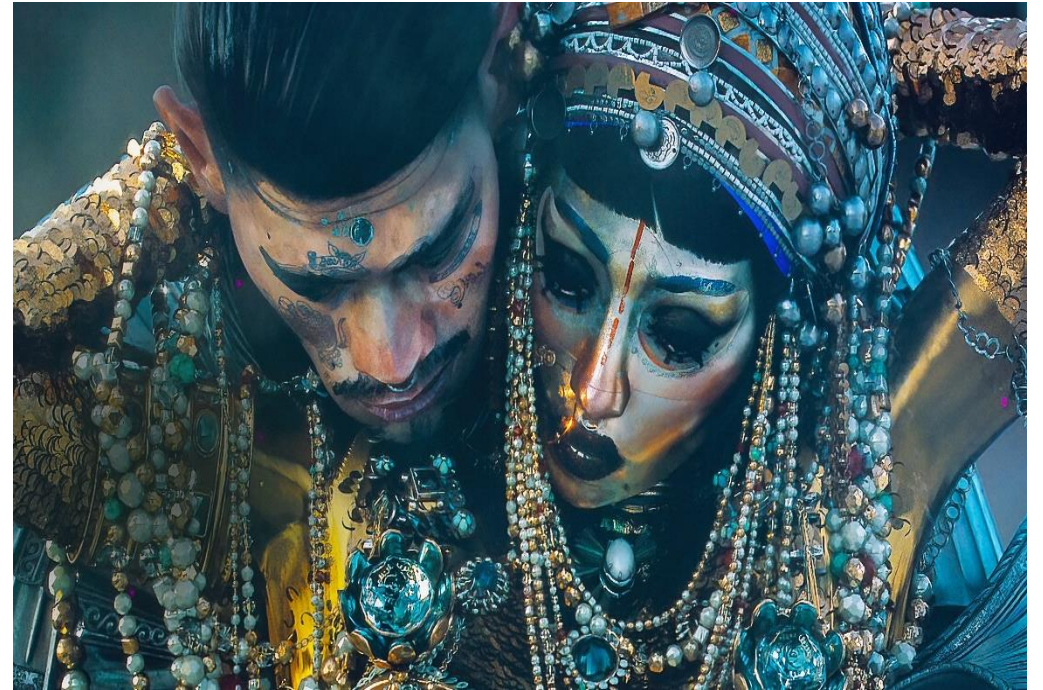
- 다양성이 높으면 적어도 리스트 내 일부 아이템이 고객에게 적합할 가능성이 높아진다.
    - 유사성 지표 등을 활용 가능

- 6. 적용 범위

- 전체 아이템 대비 추천 서비스가 추천할 수 있는 아이템의 비율을 측정
    - 카탈로그 적용 범위: 추천 결과가 고객마다 거의 유사할 가능성을 최소화하는 것이 목표
      - 적어도 한 가지 추천 리스트에 등장하는 아이템의 비율 등으로 측정

# 추천 (Recommendation)

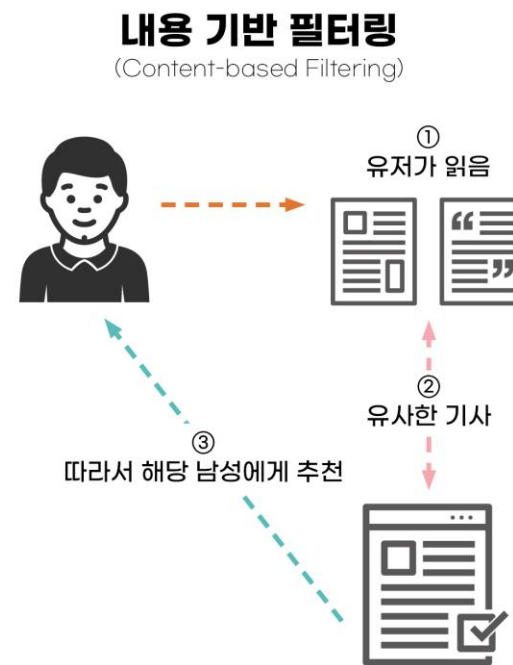
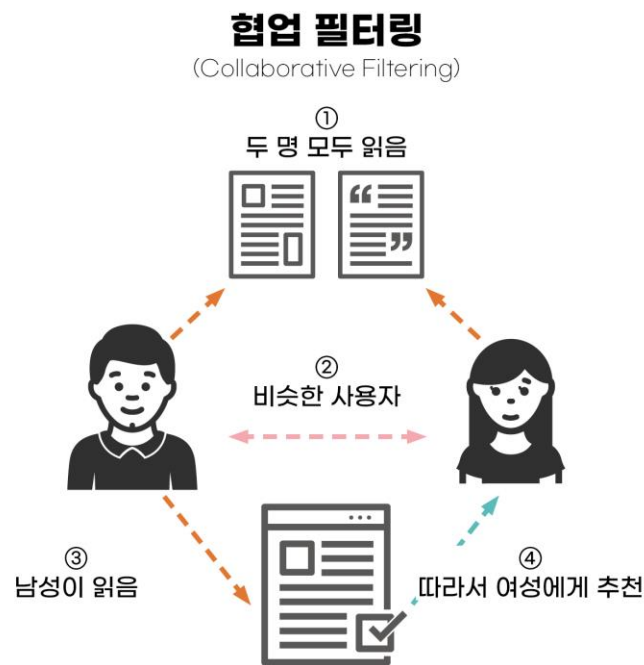
- 논의
  - 추천 시스템의 정확도가 높을수록 무조건 좋은 것인가?
  - 사례: Love, Death and Robots (2019~)
    - “The ORDER OF THE EPISODES for Netflix’s new series Love Death & Robots changes based on whether Netflix thinks you’re gay or straight.”



# 추천 (Recommendation)

## ■ 추천 기법의 개관

- 콘텐츠 기반 필터링: 아이템의 콘텐츠 데이터에 의존
- 협업 필터링: 고객 평점 행렬 내의 패턴에 의존
- 하이브리드: 핵심 알고리즘들의 결합
- 맥락 기반 추천: 시간, 장소, 마케팅 채널 등과 같은 맥락적 정보까지 반영



# 추천 (Recommendation)

- 콘텐츠 기반 필터링

- 사용자가 과거에 긍정적으로 평가했던 아이템을 선택하고 그 아이템과 비슷한 다른 아이템을 추천
- 콘텐츠와 카탈로그 아이템 내용 사이의 유사성을 측정
  - 고객 간의 구매/평가 지표의 유사성이나 고객의 행동 데이터 등을 포함하지 않음.
  - 추천이 제안되는 고객의 프로파일만 사용

- 주요 방법론

- k-NN 알고리즘, 나이브 베이즈 분류 모형 등
- Vector space 기반 모형 또는 잠재 토픽 모형(예: LSA, LDA 등)



# 추천 (Recommendation)

- 콘텐츠 기반 필터링의 장점

- 1. 사용자 데이터로부터의 독립성
  - 전체 사용자 수가 적거나 수집된 전체 평점 수가 적을 경우에 매우 유용
  - 독특한 취향을 가진 고객에게도 좋은 추천 서비스를 제공 가능
- 2. 새롭고 희귀한 아이템
  - 고객 평점이 거의 없는 새로운 또는 희귀한 제품의 추천이 가능
  - 롱테일(long tail)의 맥락에서 볼 때 매우 중요한 장점

# 추천 (Recommendation)

- 콘텐츠 기반 필터링의 장점

- 3. 다른 카테고리의 추천

- 고객 구매 패턴의 유사성만 가지고 추천한다면 다른 카테고리를 추천하는 결과가 잘 안 나올 가능성 있음. → 특히 여러 카테고리에 걸쳐 구매를 한 고객이 적을 경우

- 4. 해석 가능성

- 추천 결과를 고객에게 설명하기에 유용
    - 예: 영화 추천의 경우, “A라는 일본 액션 영화를 좋아하셨기 때문에 B라는 일본 액션 영화를 추천합니다.”

# 추천 (Recommendation)

- 콘텐츠 기반 필터링의 단점

- 1. 콘텐츠 특징(feature) 추출의 문제
  - 유사성 측정의 대상이 되는 콘텐츠 특징 추출 결과의 품질은?
  - “인간의 취향”을 제품의 기본적인 속성 등으로 쉽게 표현할 수 있는가?
  - 콘텐츠를 구성하는 텍스트, 이미지 등으로부터 특성을 적절하게 추출할 수 있는가?
- 2. 신규 사용자 추천
  - 과거 데이터가 없는 신규 사용자에게 대한 추천 서비스가 어려움.
- 3. 사소한 추천
  - 참신하지 않은, 또는 우연적이지 않은 추천을 할 가능성이 상대적으로 높음.

# 추천 (Recommendation)

- 협업 필터링

- 다른 사용자들의 피드백에 기반을 둔 추천 시스템
  - “인간의 취향”을 고객의 평점 데이터를 이용해 측정
  - 관측하기 어려운 고객의 취향, 심리, 판단 등에 대한 많은 정보를 획득할 수 있는 시스템
- 원래 다른 사람의 피드백을 바탕으로 한 이메일 필터링 시스템으로 시작
  - 다른 사람의 피드백으로부터 정보를 획득한다는 아이디어가 주목받으면서 아마존, 넷플릭스 등 대규모 산업적 추천 시스템에 응용되기 시작
- 핵심은 고객 평점 데이터의 완성: 즉, 고객 평점의 예측 모형화가 중요!

# 추천 (Recommendation)

- 협업 필터링

- 이웃 기반 협업 필터링

- 가정: 비슷한 사용자들은 비슷한 취향과 선호를 가진다
    - 비슷한 다른 고객들의 과거 평점을 이용해 특정 고객의 미래 평점을 예측
      - 예: 이웃 사용자들의 과거 평점들의 평균
  - 단계 1: 누구를 이웃(비슷한 사용자)으로 정의하여야 하는가?
  - 단계 2: 이웃의 평점 평균을 이용해 특정 고객의 평점을 어떻게 예측할 것인가?

# 추천 (Recommendation)

- 협업 필터링

- 아이템 기반 협업 필터링

- 다른 고객들의 평점에 기반을 두고 아이템 간 유사성 지표를 계산하여 특정 고객에 의해 긍정적으로 평가된 아이템과 비슷한 아이템을 추천
    - “이웃” 개념을 활용한다는 점에서 콘텐츠 기반 추천 시스템과는 차이

- 모델 기반 협업 필터링

- “이웃”에만 주목할 것이 아니라 고급 기법을 사용해 예측 모델을 구현
    - 회귀 분석 모형, 결측치 예측 모형 등 다양한 모형 활용 가능

# 추천 (Recommendation)

- 협업 필터링의 장점

- 1. 아이템 콘텐츠의 추가 정보 없이도 추천 시스템 구축 가능
- 2. 고객의 암묵적/심리적 프로파일의 활용 가능
  - 고객 행동 패턴으로부터 고객의 기호와 판단을 유추 가능
- 3. 사소하지 않은 추천의 가능
  - 관측되지 않은 고객의 프로파일을 활용하기 때문에 참신한 혹은 우연적인 추천이 가능

# 추천 (Recommendation)

- 협업 필터링의 단점

- 1. 과거 데이터가 부족할 경우
  - 상당한 양의 신뢰할 만한 고객 평점 데이터가 구축된 상황에서만 제공 가능
- 2. 새로운 사용자와 아이템의 경우
  - 과거 평점 기록이 없는 새로운 사용자나 새롭게 제공되는 아이템의 경우 예측이 어려움.
- 3. 인기도 편향 / 추천 제품의 표준화
  - 평점 데이터의 패턴에 기반을 두고 추천하기 때문에 인기 있는 아이템이나 ‘대중적인’ 아이템이 추천될 가능성이 높아짐.



# 추천 (Recommendation)

- 하이브리드 추천 기법

- 스위칭 (switching): 조건에 따라 알고리즘을 바꾸는 형태로 여러 알고리즘을 결합
- 블렌딩 (blending): 여러 개의 알고리즘으로 도출된 평점 예측치를 특정 규칙에 따라 결합

- 맥락 기반 추천

- 고객의 위치, 시간, 의도, 마케팅 채널, 기타 여러 조건 정보를 참고하여 추천 서비스 제공
- 맥락 사전 필터링, 맥락 사후 필터링 등

# 추천 (Recommendation)

- Recommendation system in Marketing
- A. Ansari, S. Essegaier, and R. Kohli (2000), Internet recommendation systems, *Journal of marketing research*, 37(3), 363-375.
  - We use a hierarchical Bayesian approach to design a recommendation system.
    - To allow unobserved heterogeneity in consumer preferences
    - To assess the effect of unobserved product heterogeneity on preferences to allow for the introduction of unobserved product attributes
  - We suggest that preference models can offer good alternatives and this approach allows statistical integration of five types of information useful for making recommendations: a person's expressed preferences, preferences of other consumers, expert evaluations, item characteristics, and individual characteristics.

# 추천 (Recommendation)

- Recommendation system in Marketing
- A. Ansari, S. Essegai, and R. Kohli (2000), Internet recommendation systems, *Journal of marketing research*, 37(3), 363-375.
  - Customer heterogeneity

The observations for each customer are used to specify a customer-level regression model:

$$(1) \quad r_{ij} = \mathbf{w}_j \boldsymbol{\beta}_i + e_{ij}, e_{ij} \sim N(0, \sigma^2),$$

where  $j \in M_i$ ,  $\mathbf{w}_j$  is a vector of movie attributes (genre and expert ratings) for movie  $j$ , and  $\boldsymbol{\beta}_i$  is a vector of parameters that represent the preference structure for customer  $i$ .

$$(2) \quad \boldsymbol{\beta}_i = \mathbf{z}_i \boldsymbol{\mu} + \boldsymbol{\lambda}_i,$$

for  $i = 1$  to  $I$ . In Equation 2,  $\mathbf{z}_i$  contains the characteristics of customer  $i$ , and  $\boldsymbol{\lambda}_i$  represents the unobserved customer effect for the  $i$ th customer.

# 추천 (Recommendation)

- Recommendation system in Marketing
- A. Ansari, S. Essegai, and R. Kohli (2000), Internet recommendation systems, *Journal of marketing research*, 37(3), 363-375.
  - Product heterogeneity

Let  $C_j = \{i_1, i_2, \dots, i_{n_j}\}$  represent the index set of the  $n_j$  customers who rated movie  $j$ . Let  $r_{ji}$  represent the rating given by customer  $i$  for movie  $j$ , where  $i \in C_j$ . The number of customers that provide ratings for a movie varies, which yields an unbalanced data set. The observations for movie  $j$  can be used in specifying a movie-level regression model as follows:

$$(4) \quad r_{ji} = \mathbf{z}_i' \boldsymbol{\beta}_j + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2),$$

$$(5) \quad \boldsymbol{\beta}_j = \mathbf{w}_j' \boldsymbol{\mu} + \boldsymbol{\gamma}_j, \quad \boldsymbol{\gamma}_j \sim N(\mathbf{0}, \boldsymbol{\Gamma}),$$

for  $j = 1$  to  $J$ . The vector  $\mathbf{w}_j$  contains the observed movie characteristics, and  $\boldsymbol{\gamma}_j$  represents the unobserved movie effects. The complete model can alternatively be written as

Table 2  
MODEL COMPARISON STATISTICS

<i>Models</i>		<i>Log-Marginal Likelihood</i>	<i>DIC Statistics</i>		
<i>Heterogeneity</i>	<i>Movie Attributes</i>		<i>Fit D</i>	<i>Complexity pD</i>	<i>DIC</i>
No heterogeneity	Genre only	-18,801	37,589	13	37,602
	Expert only	-18,398	36,788	8	36,796
	Genre and expert	-18,327	36,638	17	36,655
Customer heterogeneity	Genre only	-17,581	34,135	1020	35,155
	Expert only	-17,162	33,429	900	34,329
	Genre and expert	-16,909	32,215	1501	33,716
Movie heterogeneity	Genre only	-18,072	35,825	275	36,100
	Expert only	-18,067	35,834	259	36,093
	Genre and expert	-18,066	35,830	260	36,090
Movie and customer Heterogeneity	Genre only	-16,793	32,118	1390	33,508
	Expert only	-16,840	32,502	1146	33,648
	Genre and expert	-16,675	31,488	1717	33,205

Notes: All models include demographic variables.

# 추천 (Recommendation)

- **Recommendation system in Marketing**
- **Ansari, A., and Mela, C. F. (2003). E-customization. *Journal of marketing research*, 40(2), 131-145.**
  - Few models exist to help firms implement one-to-one marketing on the Internet.
  - We develop a statistical and optimization approach for customization of information on the Internet.
  - Specifically, we develop an approach that enables websites to customize permission-based e-mail communications to increase website traffic (though the approach can be more broadly applied to the issue of website customization).