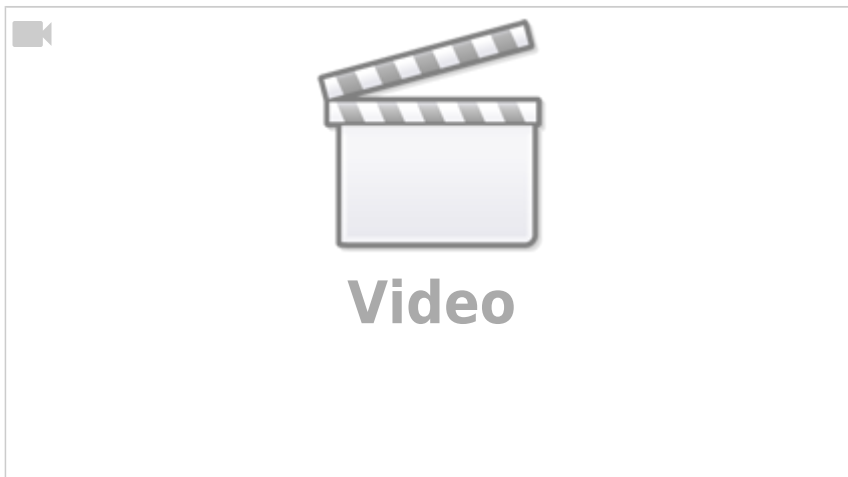


추천 시스템

추천 시스템의 기본 유형

- 콘텐츠 기반 필터링 [Content Based Filtering](#)
- 협업 필터링 [Collaborative Filtering](#)
 - 최근접 이웃 기반 협업 필터링 [Nearest neighbor collaborative filtering](#)
 - 잠재 요인 협업 필터링 [Latent factor collaborative filtering](#)



Statistics-based

좋아할 만한 = 통계적으로 유의미한 아이템

Chi-Squared (예상보다 많이 본) - Categorical Cross-Entropy(KL-divergence) (분포가 많이 다른) - Continuous

Chi-Squared 개념과 활용

$$\chi^2 = \sum_{i=1}^k \frac{(x_i - m_i)^2}{m_i} \quad x = \text{관측치} \quad m = \text{예측치}$$

Categorical Variable

유저가 소비한 아이템의 예측치와 실제로 소비한 관측치의 차이를 이용

예상보다 많이 본 것이므로, 절대값보다는 상대적인 변화량에 주목

Collaborative Filtering

Co-occurrence 동시에 발생하는 이벤트에 주목하라

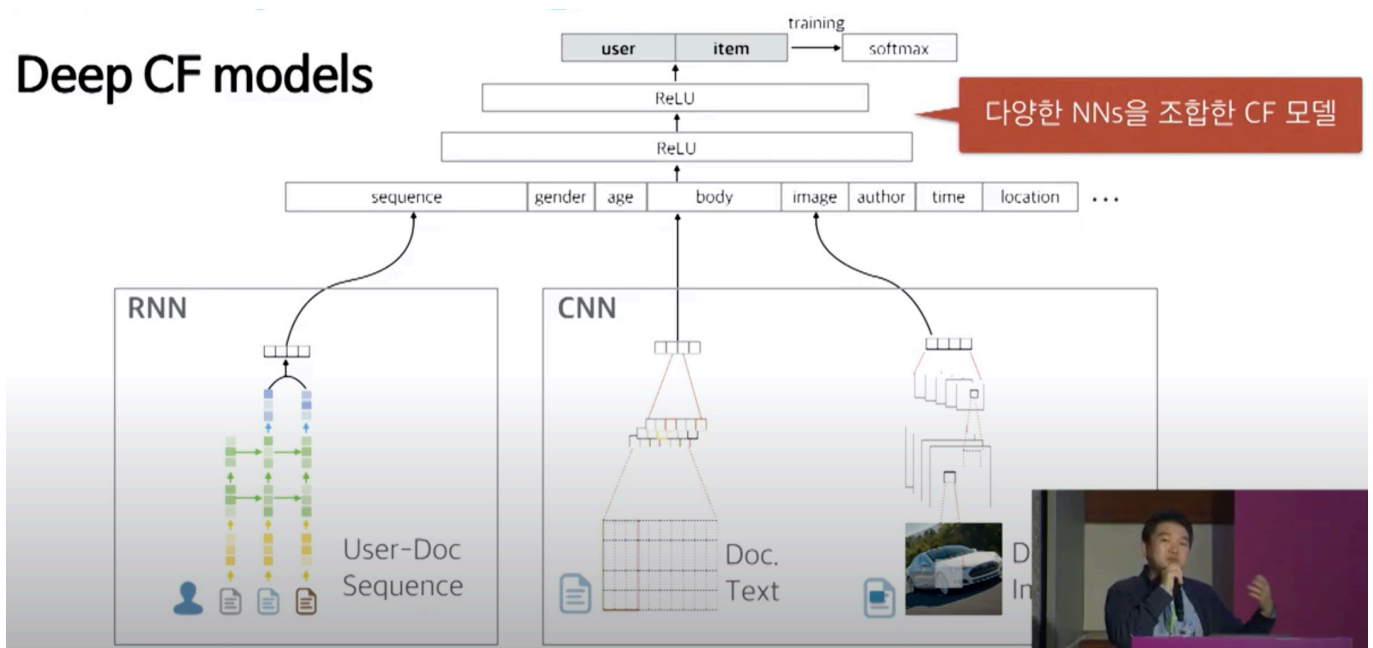
PMI (Pointwise Mutual Information) $PMI(A, B) = \frac{P(A, B)}{P(A)P(B)}$ 함께 발생한 빈도와 함께 각 이벤트가 발생할 확률을 함께 고려한 정보량

Deep Learning

RNN for News Recommendation

유저의 문서 소비 패턴을 보고 추천

Deep CF models



추천 품질 평가 요소

추천의 만족도

- Accuracy: 유저가 실제 소비한 아이템이 상위에 추천되는지
- Diversity: 다양한 주제/유형의 아이템이 잘 추천되는지
- Novelty: 새로 나온 최신의 아이템이 잘 추천되는지

Agile Development

Language & Tool 통합 데이터 분석 및 추천 모델링에 사용하는 언어와 도구를 실서비스 투입이 가능하도록 통합

- [Airflow](#) (scheduler)
- [Hive](#) (data warehouse)
- [Spark](#) (big data processing engine)
- [Hadoop YARN](#) (distributed cluster resource manager)
- [Slider](#) (an application to deploy existing distributed applications on Yarn cluster)
- [OpenTSDB/Grafana](#) (a scalabe, distributed monitoring system)
- [DOT](#) (distributed incremental search engine)
- [DDK/Cana](#) (event-driven near-realtime serverless compute solution)

- [C3](#) (PassS, Hadoop cluster)
- [Cuve](#) (PassS, HBase, Kafka)
- [nBase-ARC](#) (PaaS)

링크

- <https://yamalab.tistory.com/67>

관련 문서

Plugin Backlinks: 아무 것도 없습니다.

From:
<https://moro.kr/> - Various Ways

Permanent link:
<https://moro.kr/open/%EC%B6%94%EC%B2%9C-%EC%8B%9C%EC%8A%A4%ED%85%9C>

Last update: 2020/11/02 09:03

