2021-7-22 Seminar

Privacy-Preserving News Recommendation Model Learning

Findings of ACL: EMNLP 2020

발표자:배수영

Content

1. Introduction
2. Related Work
3. FedNewsRec Framework
4. Experiment

Introduction

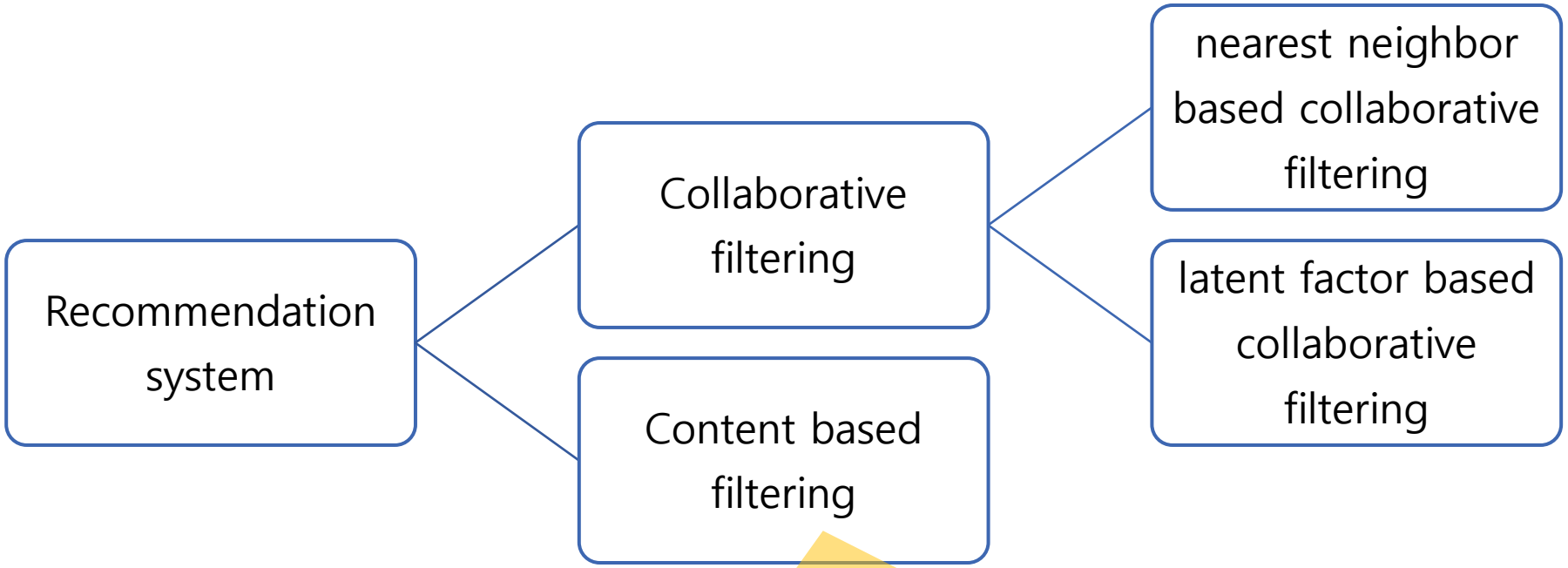
- 인터넷 발달로 모바일&웹 상 신문 기사 양이 방대하게 증가.
- User마다 관심있는 주제가 다르기 때문에 personalized news recommendation사용.
- 기존 추천 시스템은 중앙화(centralized storage)된 저장 장치 사용.
- 사용자의 행동이 정보 유출에 예민하다.
- 데이터 학습을 서버로 모으지 않고, 각각 사용자의 device에서 지역적으로 저장하는 privacy preserving method 제안

Contribution

- Propose a privacy-preserving method
- Apply local differential privacy(LDP)
- Conduct extensive experiments on a real-world dataset

Related Work
1. News Recommendations

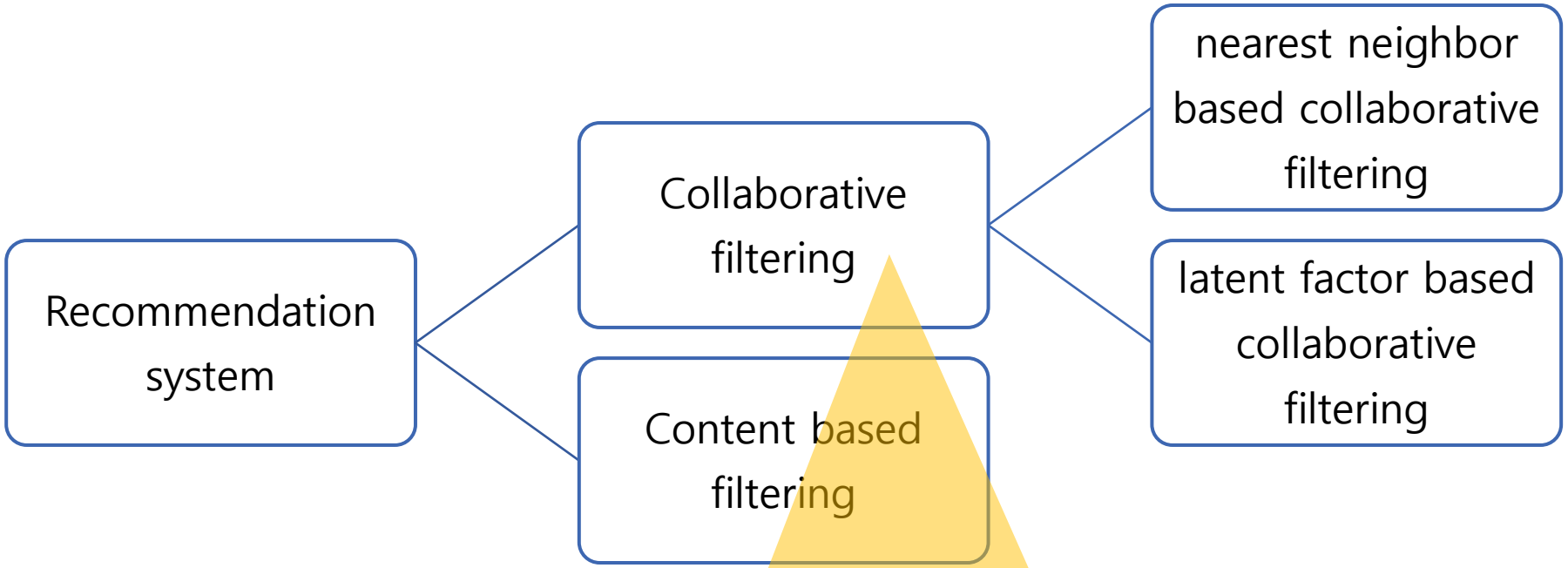
- Overal Concept of Recommendation system



사용자가 특정 아이템을 선호하는 경우 그 아이템과 비슷한 콘텐츠를 가진 다른 아이템 추천해주는 방식

Related Work
1. News Recommendations

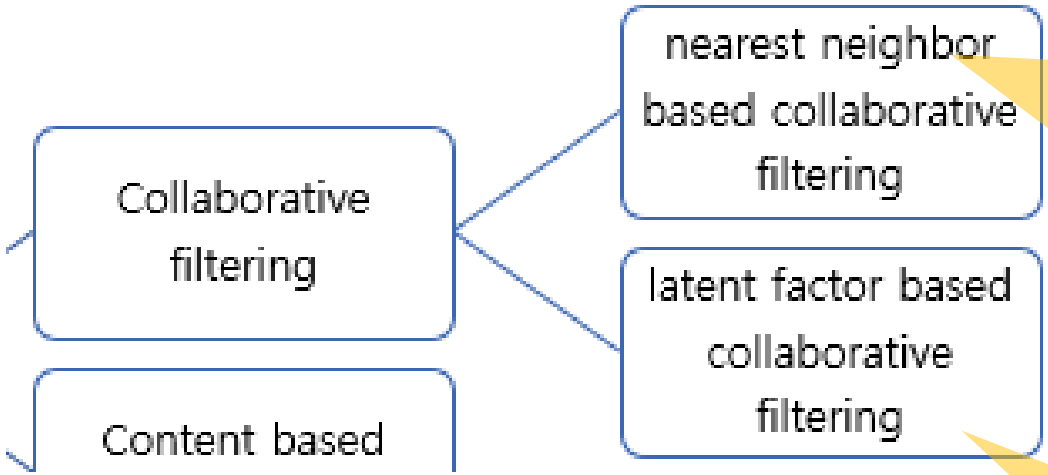
- Overal Concept of Recommendation system



사용자와 비슷한 사용자를 찾아 그룹을 형성 한 후 그룹 간 평가 점수와 선호도 고려해 추천하는 방식

Related Work
1. News Recommendations

- Overal Concept of Recommendation system



사용자-아이템 관계를 행렬로 만들어서 사용자가 아직 평가하지 않는 아이템을 예측하는 것 (사용자 기반, 아이템 기반)

사용자-잠재요인, 잠재요인-아이템 으로 행렬 분해(matrix factorization) 진행

https://lsjsj92.tistory.com/564?category=853217

Related Work
1. News Recommendations

- 세 가지 핵심 과정

1. 뉴스 기사 모델링
   - feature based method
   - denoising autoencoder
   - multi-head self-attention network
2. 사용자 취향(User Behavior) 모델링
   - GRU network
   - long-and short-term user representation model
3. 뉴스 기사와 사용자 취향 관계 연결 모델링
   - dot product of user and news representation vectors
   - cosine similarity
   - feed-forward network
   - feature-interaction network

Related Work
2. Federated Learning

- 연합학습: 다수의 로컬 클라이언트와 하나의 중앙 서버가 협력해 데이터가 탈 중앙화된 상황에서 글로벌 모델을 학습하는 기술
- 각 user device에서 locally-computed model의 gradient를 server로 update (in shared model).
- Model update시 raw user data 보다 정보를 덜 담고 있기 때문에 (gradient) 개인정보 문제 해결 가능
- 하지만 연합 학습은 real-world recommendation scenario에 적용하기에 실용적이지 않다.

Related Work
3. Local Differential Privacy

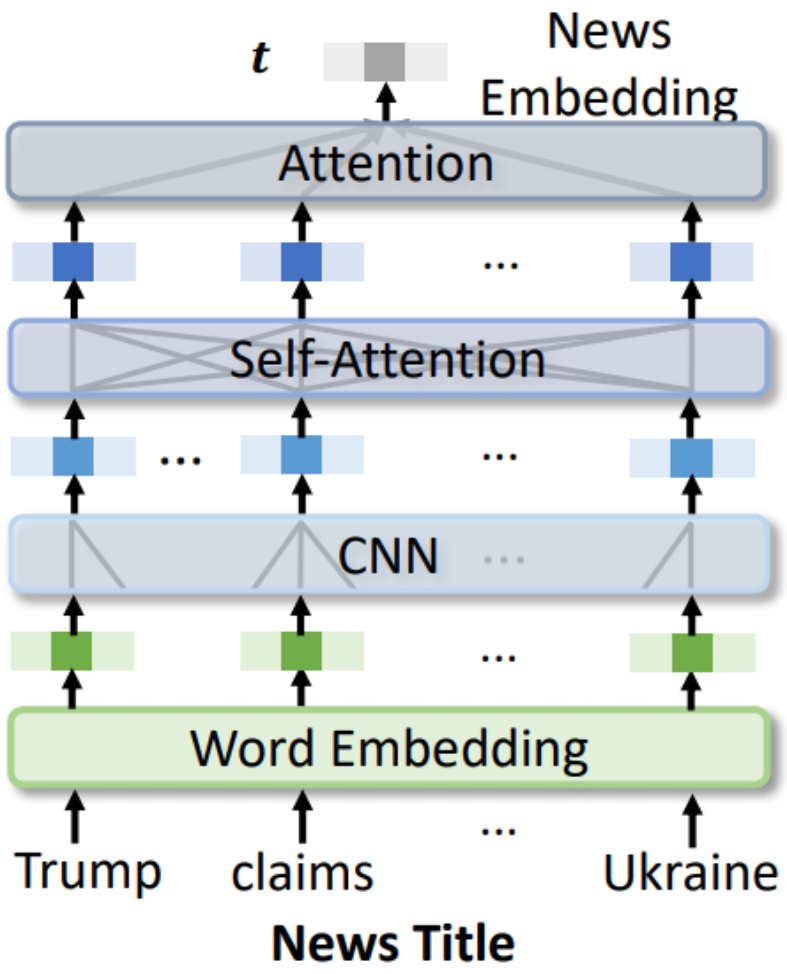
- 각 사용자들의 private value(v)를 untrusted third-party aggregator에 전달 -> private value distribution의 통계적 정보를 학습한다.
- Randomized algorithm적용

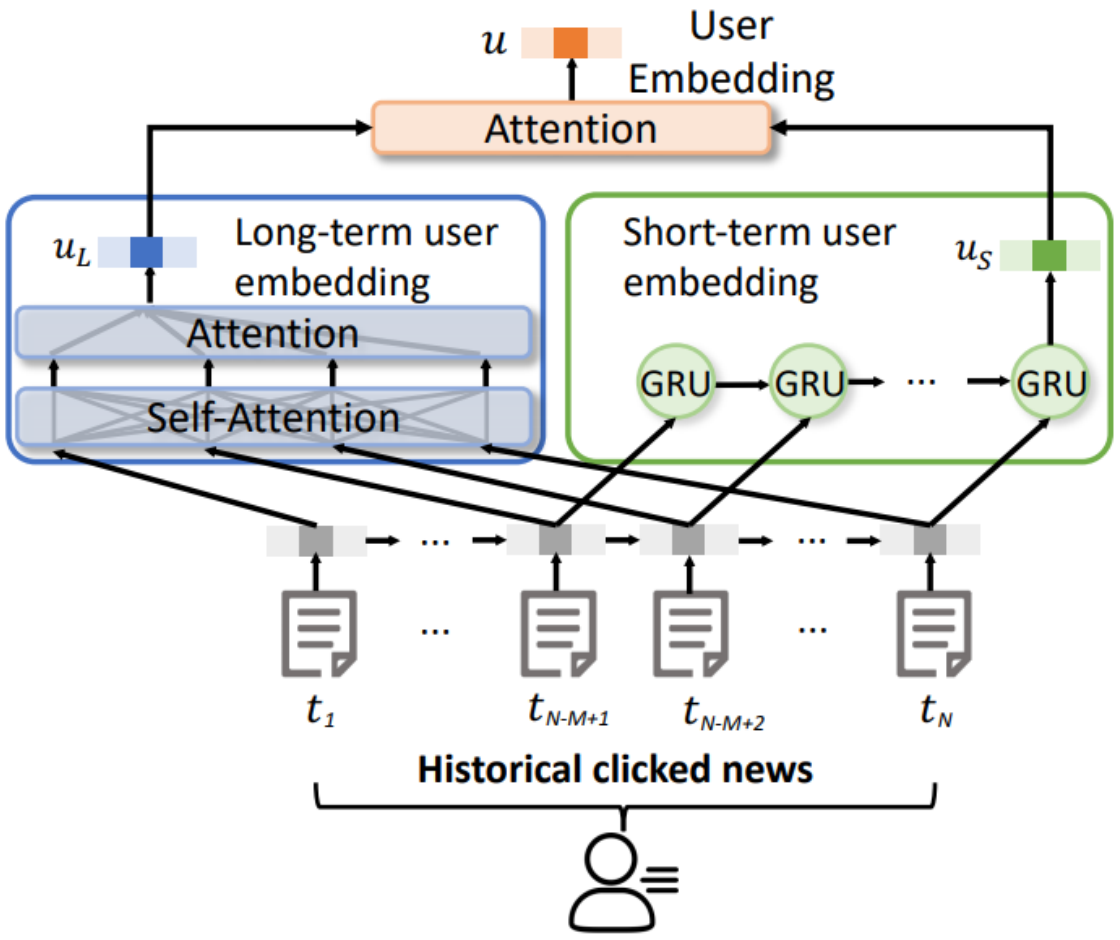
$$\Pr[\mathcal{M}(v) = y] \leq e^{\epsilon} \Pr[\mathcal{M}(v') = y],$$

- private value가 두개만 있다고 가정할 때 각각을 v, v' 라고 함
- ε (privacy budget)가 작을 수록 better private information protection
- LDP를 user behavior 기반 model gradient에 적용해 server에 업로드 한다.

FedNewsRec

- Basic News Recommendation Model
- News Model & User Model

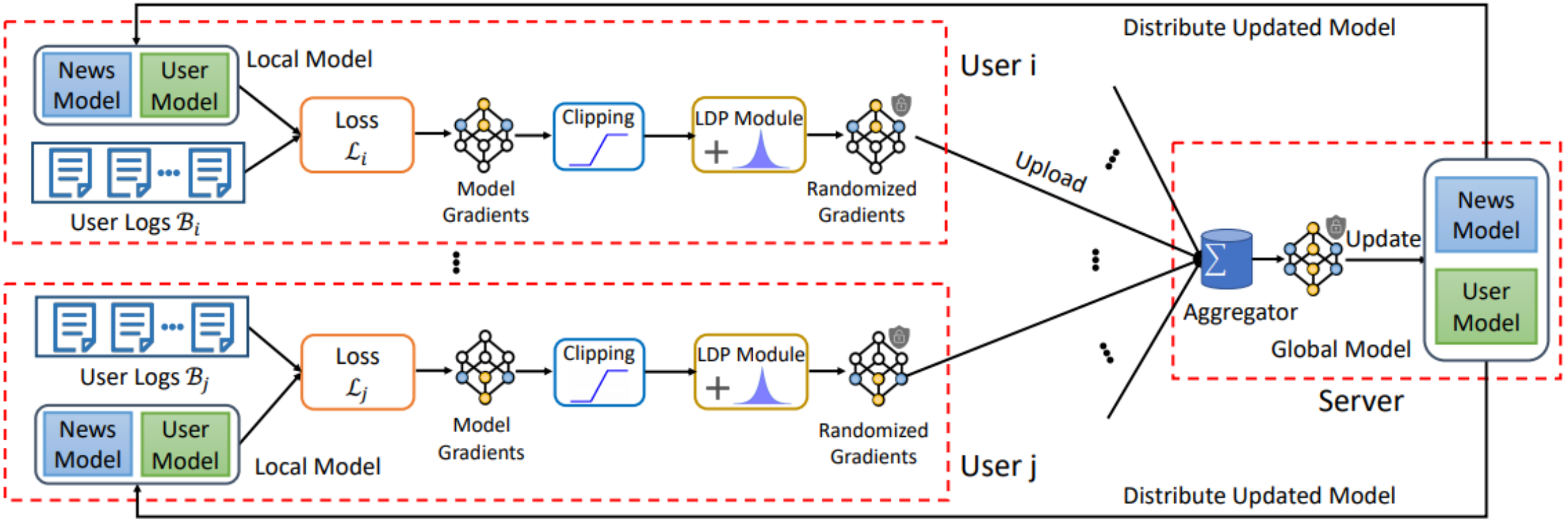




LSTUR(neural news recommendation with long and short term user representation)
https://nvagus.github.io/paper/ACL19NewsRec.pdf

FedNewsRec

- The framework for FedNewsRec



FedNewsRec

- The framework for FedNewsRec

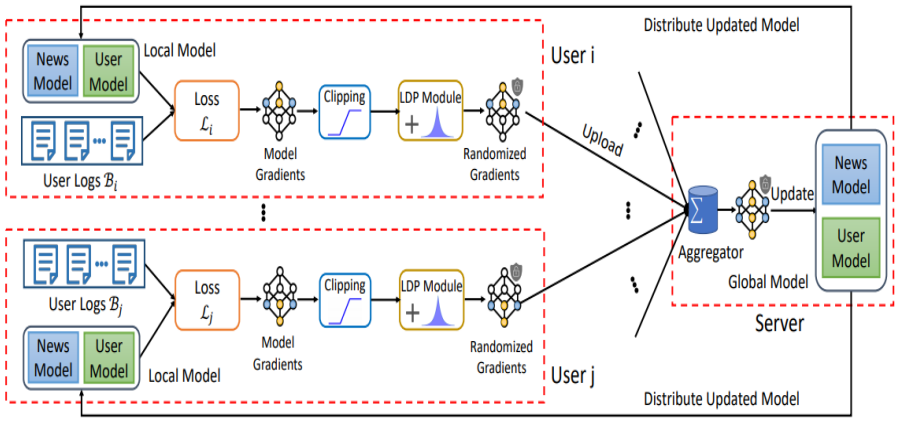
1. User behavior 정보들은 user device(client) 에서만 사용된다.
2. 각 client에는 현재 모델의 copy가 들어 있습니다.
3. Local model gradient에 Clipping, LDP 추가

$$\mathcal{M}(g_u) = \mathbf{clip}(g_u, \delta) + n,$$

Laplace noise

Laplace noise 범위를 제한해주는 parameter

$$n \sim La(0, \lambda),$$



4. Randomized local model gradient 중에서 무작위로 r% user 선택해서 update

$$\bar{g} = \frac{1}{\sum_{u \in \mathcal{U}} |\mathcal{B}_u|} \sum_{u \in \mathcal{U}} |\mathcal{B}_u| \cdot \tilde{g}_u,$$

Aggregated model gradients

$$\Theta = \Theta - \eta \cdot \bar{g},$$

Updated global model parameter

FedNewsRec

- 왜 FedNewsRec 구조가 사용자 정보를 보호할 수 있나?

1. User behavior data가 user own device에 저장되고 server로 전달 안 된다. Gradient가 전달된다.
2. Local model gradients가 group behavior로부터 계산된다. (not single behavior) 따라서 서버에서 특정 행동을 추론하기 어렵다.
3. Laplace noise추가해서 local differential privacy technique 적용한다. 개인 정보를 강화한다.

Experiment

- Dataset and Experimented Settings

	MSN-News	*Adressa*
# users	100,000	528,514
# news	118,325	16,004
# impressions	1,341,853	-
# positive behaviors	2,006,289	2,411,187
# negative behaviors	48,051,601	-
avg. # title length	11.52	6.60

1. Dataset
   - Adressa: public news recommendation dataset collected from a Norwegian news website
   - MSN-News: real-world dataset collected from Microsoft News

2. Setting
   - number of self-attention head: 20
   - output dimension: 20
   - GRU hidden state: 400
   - H (number of randomly sample news): 4
   - r: 2%
   - η (learning rate): 0.5
   - δ (clip): 0.005
   - λ (laplace noise): 0.015

Experiment

- Effectiveness Evaluation
  - We compared with many methods, including:

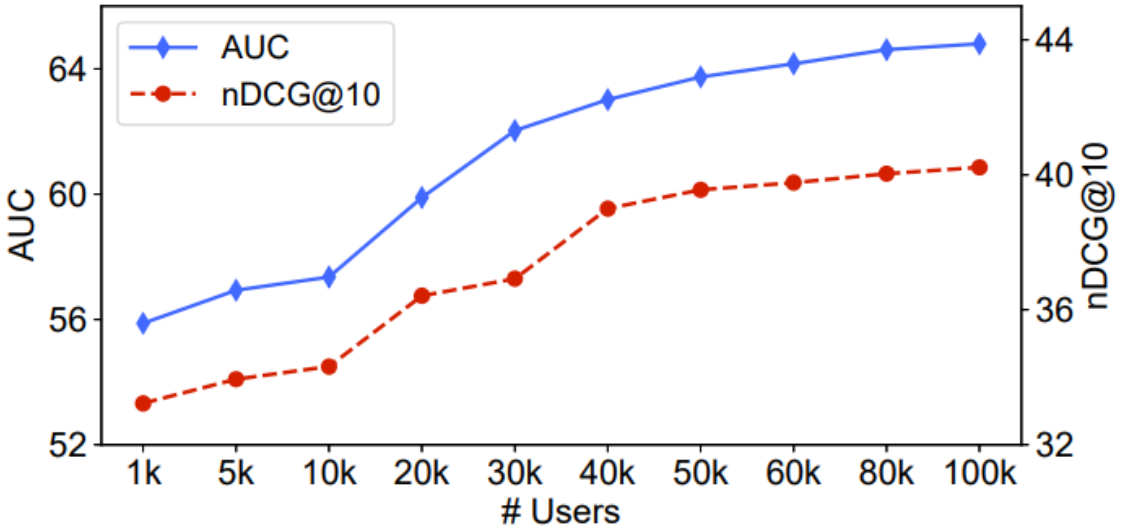
Method	MSN-News				Adressa			
	AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
FM	58.41±0.04	27.19±0.05	28.98±0.04	34.57±0.06	61.94±0.80	26.59±0.33	22.69±0.54	32.17±0.46
DFM	61.25±0.26	28.68±0.10	30.62±0.21	36.38±0.23	65.14±0.69	34.74±0.89	33.17±1.46	39.79±1.08
EBNR	63.64±0.15	29.50±0.14	31.57±0.13	37.38±0.20	65.70±0.72	30.23±0.49	29.37±0.53	36.38±0.44
DKN	62.38±0.19	29.40±0.15	31.59±0.11	37.27±0.21	67.53±1.90	32.33±2.79	31.84±2.78	39.96±2.52
DAN	62.54±0.23	29.44±0.18	31.67±0.14	37.31±0.25	64.03±3.10	33.37±2.63	31.61±3.03	38.60±3.02
NAML	64.52±0.24	30.93±0.17	33.39±0.16	39.07±0.19	69.20±2.07	35.18±1.49	34.78±1.85	42.34±1.97
NPA	64.29±0.20	30.63±0.15	33.11±0.17	38.89±0.23	66.70±2.42	34.68±1.77	33.72±2.09	41.18±1.99
NRMS	65.72±0.16	31.85±0.20	34.59±0.18	40.25±0.17	67.97±2.23	33.16±2.54	32.37±3.59	40.41±2.82
FCF	51.03±0.27	22.24±0.14	22.97±0.21	28.44±0.23	53.33±1.28	23.04±2.68	20.24±2.77	27.09±2.61
FedNewsRec	64.65±0.15	30.60±0.09	33.03±0.11	38.77±0.10	69.91±2.53	35.55±1.85	33.74±2.45	41.47±2.78
CenNewsRec	66.45±0.17	31.91±0.22	34.62±0.18	40.33± 0.24	71.02±2.09	36.31±2.52	35.73±3.71	43.98±2.52

FedNewsRec 보다 좋은 성능

FedNewsRec 보다 안 좋은 성능: centralize가 모델 훈련에 더 효과적이다.
LDP 적용한 것이 정확도를 떨어뜨린다.

Experiment

- Influence of User Number

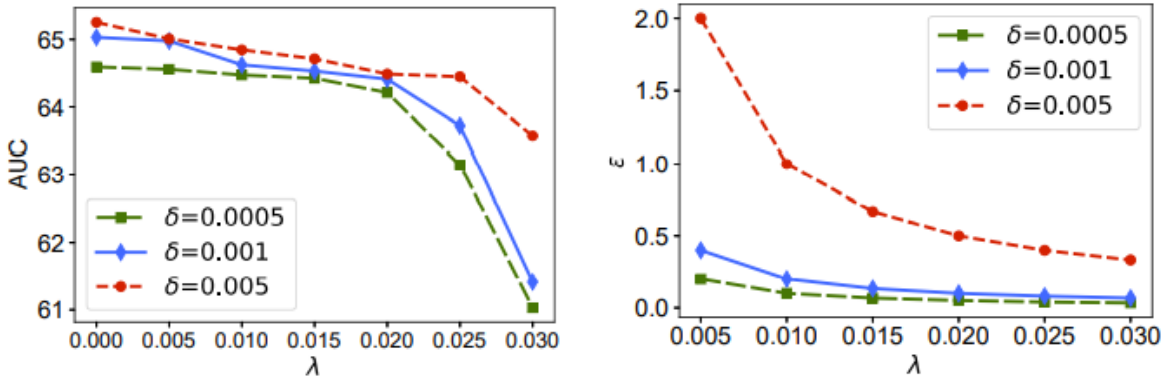


$$\mathcal{M}(g_u) = \text{clip}(g_u, \delta) + n,$$

$$n \sim La(0, \lambda),$$

$$\Pr[\mathcal{M}(v) = y] \leq e^{\epsilon} \Pr[\mathcal{M}(v') = y],$$

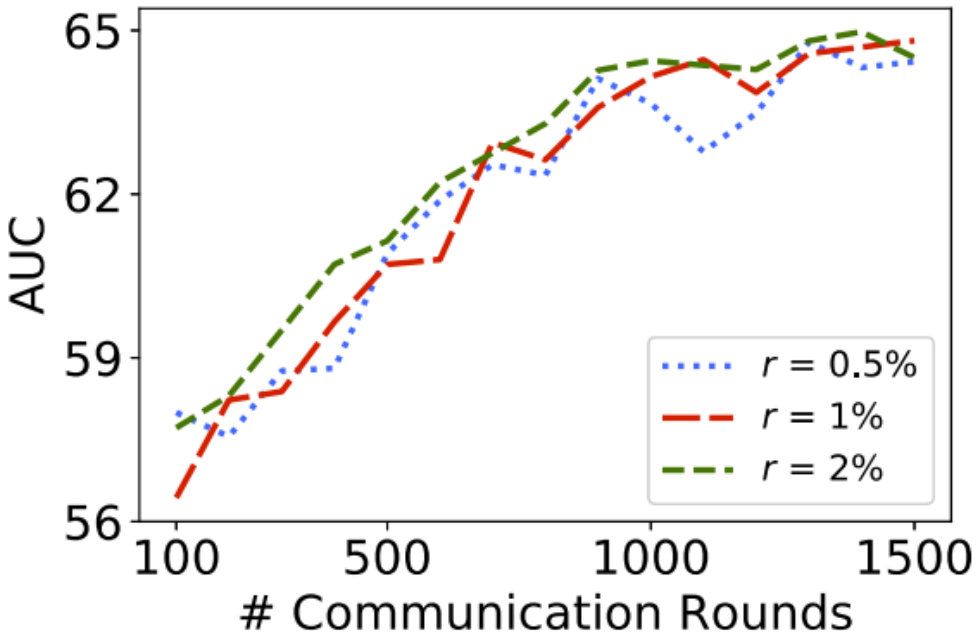
- Hyperparameter Analysis



(a) Model performance. (b) Privacy budget ϵ.

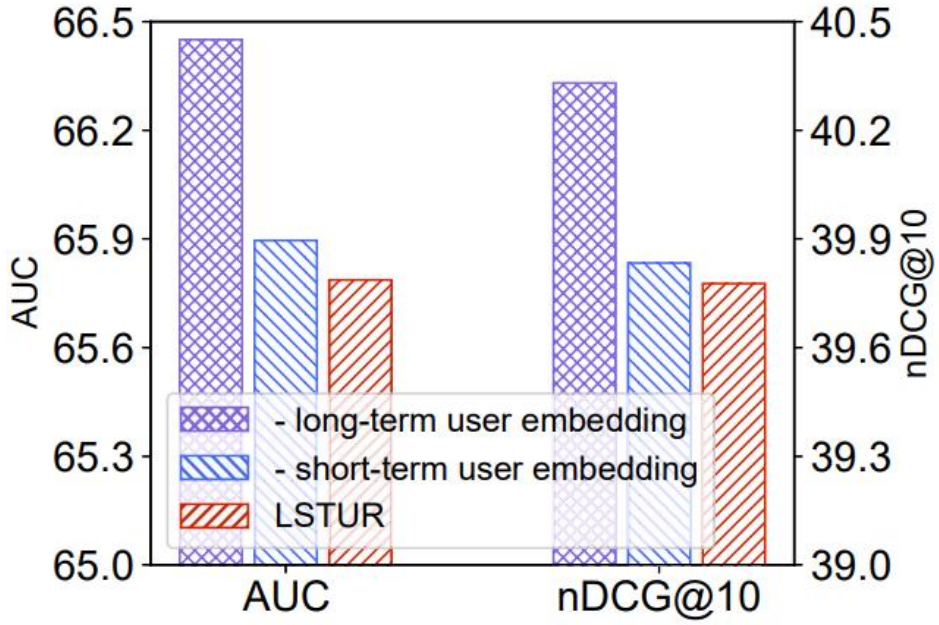
Experiment

- Convergence Analysis



약 1500 번 반복 시 효과적인 accuracy를 만족하면서 convergence

- Effectiveness of user model(ablation study)



1. Short term user embedding 제거 시 <u>성능 감소</u>
   -> 사용자들이 비슷한 뉴스를 <u>recently click</u>
2. Long term user embedding 제거 시 <u>성능 감소</u>
   -> <u>장기적인 관심사</u>에 따라 뉴스 읽는다.