Personalized News Recommendation with Candidate-aware User Modeling

Tao Qi[1], Fangzhao Wu[2], Chuhan Wu[1], Yongfeng Huang[1]

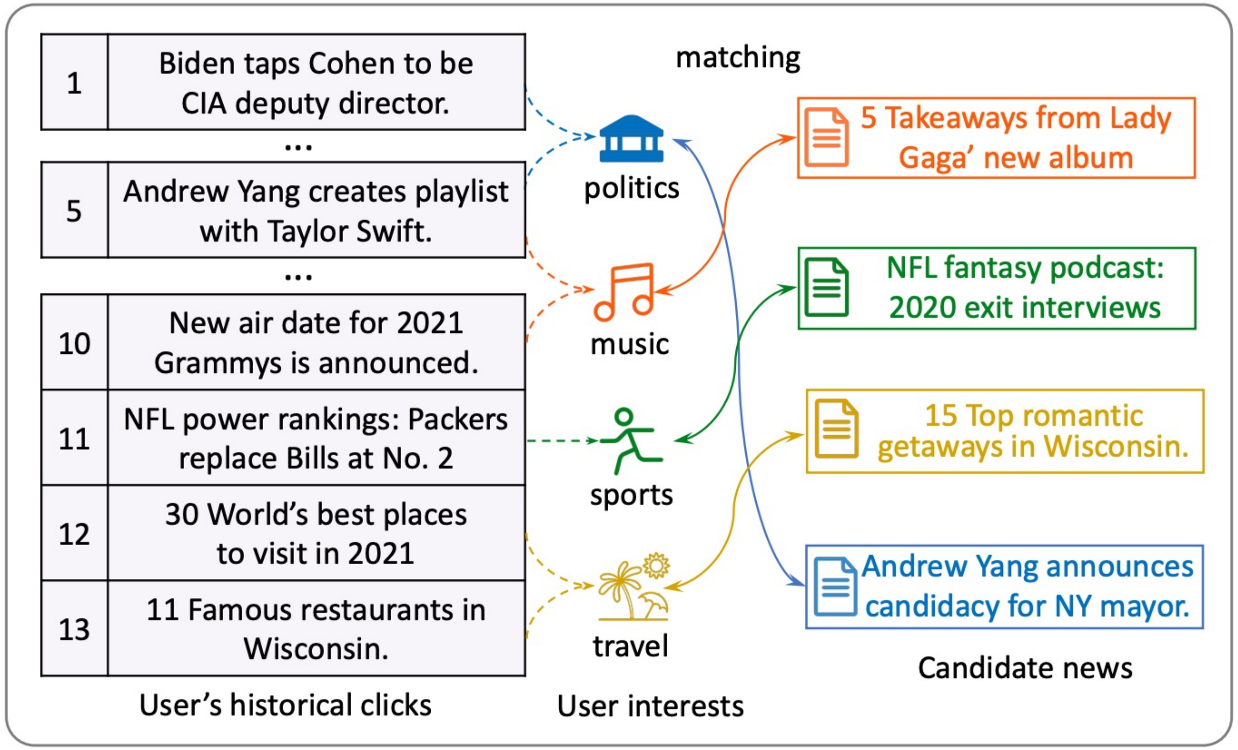
[1]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

[2]Microsoft Research Asia, Beijing 100080, China

taoqi.qt@gmail.com

User Modeling For News Recommendation

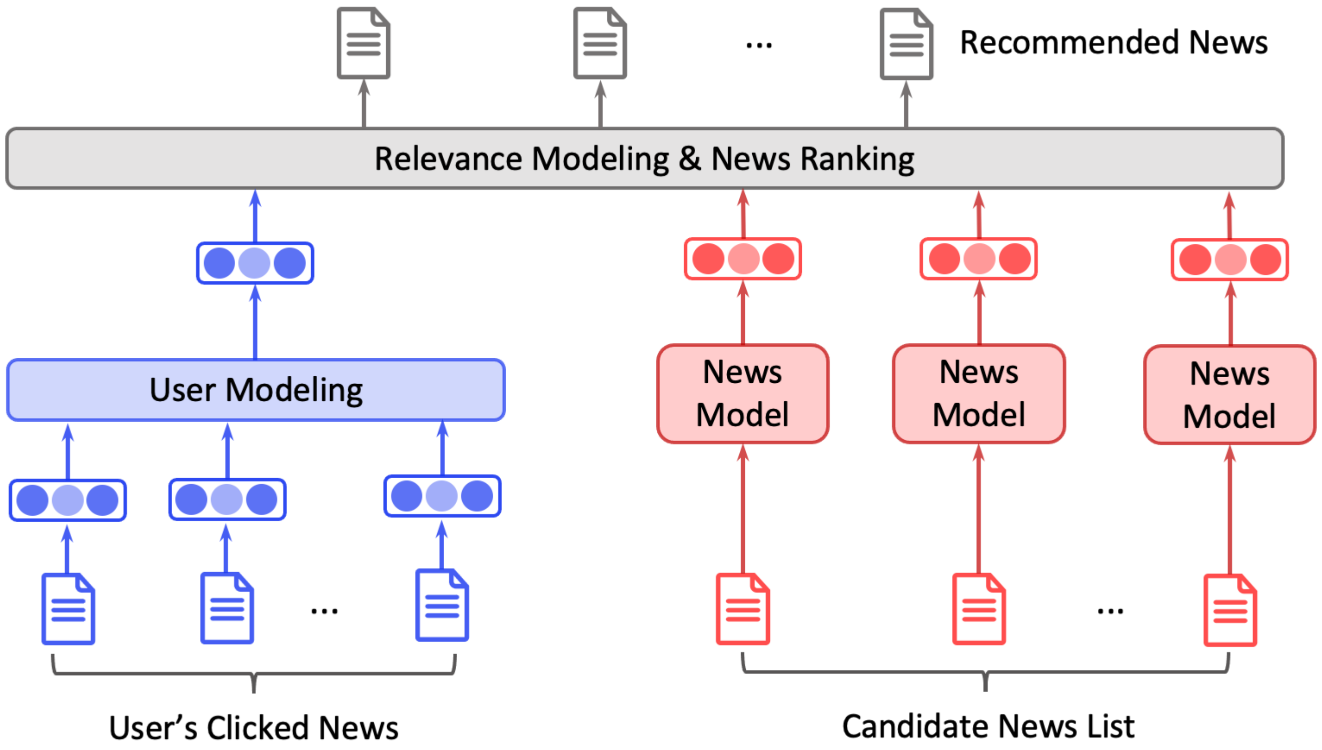
- Accurate modeling of user interest is important for news recommendation
- Users usually have multiple interest in different fields



The matching between candidate news and user interest.

User Modeling For News Recommendation

- Most existing methods model user interest in a candidate-agnostic way

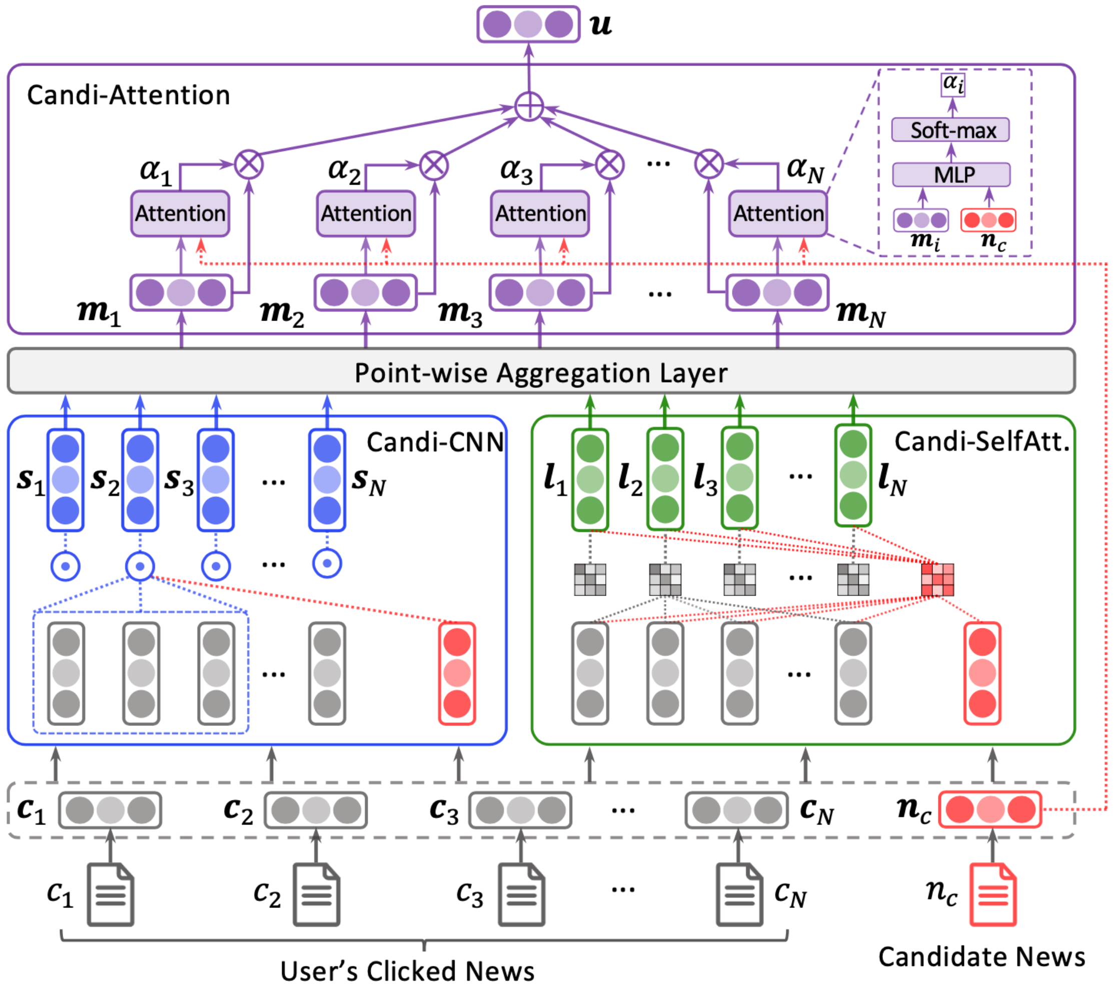


Framework of user modeling in news recommendation.

- Challenges
  - Difficult to accurately match a candidate news with a specific user interest

CAUM: Candidate-aware User Modeling

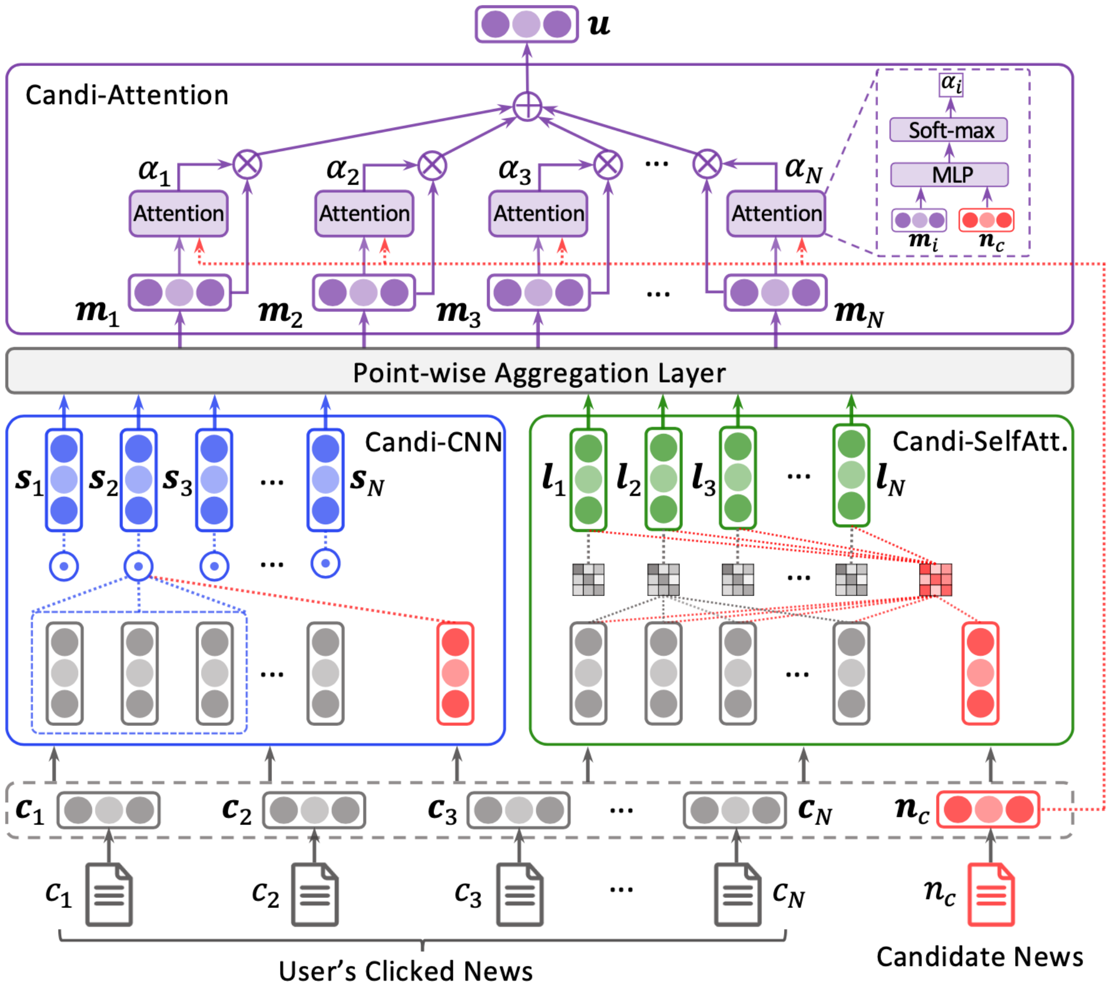
- A candidate news-aware user modeling framework for news recommendation



- Candidate-aware self-attention network
  - Long-term user interest modeling
- Candidate-aware CNN network
  - Short-term user interest modeling
- Candidate-aware attention network
  - User interest representation aggregation

Candidate-aware Self-Attention

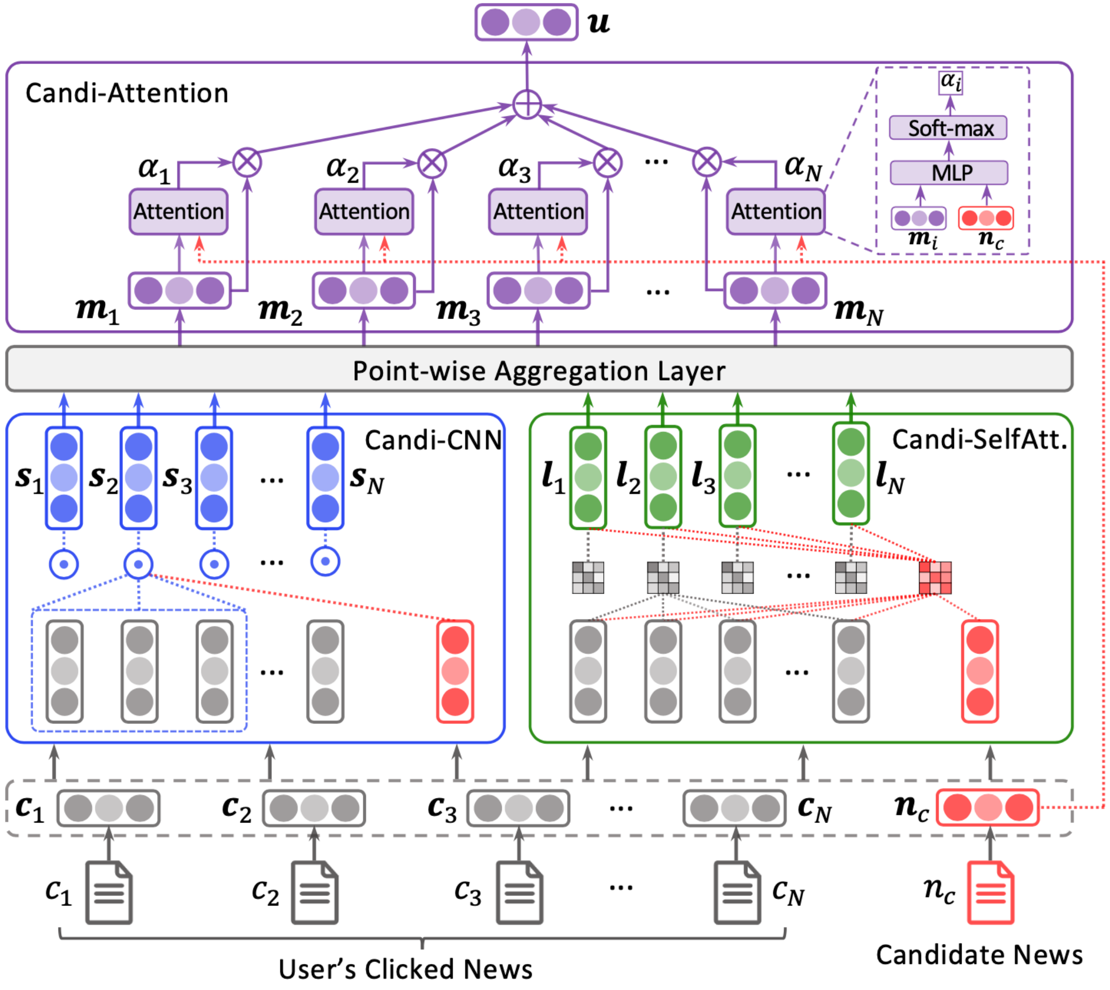
- Model long-term interest from candidate-aware global behavior contexts



- Global behavior contexts modeling
  - $\hat{r}_{i,j}^{k} = \boldsymbol{q}_i^T \boldsymbol{W}_k^r \boldsymbol{c}_j, \quad \boldsymbol{q}_i = \boldsymbol{Q}_u \boldsymbol{c}_i$
- Candidate news information fusion
  - $r_{i,j}^{k} = \hat{r}_{i,j}^{k} + \boldsymbol{q}_c^T \boldsymbol{W}_k^r \boldsymbol{c}_j, \quad \boldsymbol{q}_c = \boldsymbol{Q}_c \boldsymbol{n}^c$
- Build global contextual click representations
  - $\boldsymbol{l}_i^k = \boldsymbol{W}_o^k \sum_{j=1}^{N} \gamma_j^k \boldsymbol{c}_j,$
  - $\gamma_j^k = \exp(r_{i,j}^k) \, / \sum_{p=0}^{N} \exp(r_{i,p}^k) \, ,$
  - $\boldsymbol{l}_i = [\boldsymbol{l}_i^1, \boldsymbol{l}_i^2 \dots, \boldsymbol{l}_i^K],$

Candidate-aware CNN Network

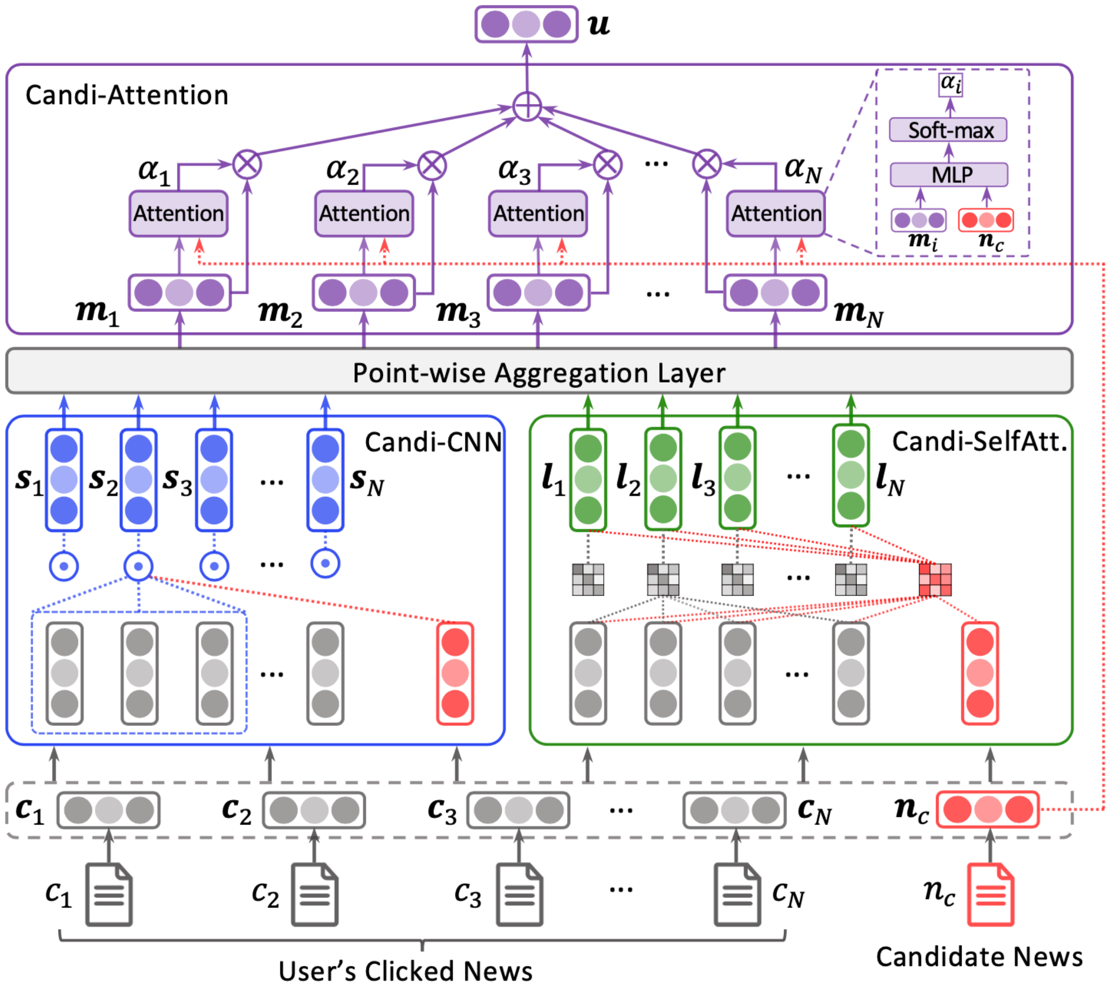
- Model short-term interest from candidate-aware local behavior contexts



- Model relatedness between candidate news and local contexts via CNN filters
- $\boldsymbol{s}_i = \boldsymbol{W}_c[\boldsymbol{c}_{i-w}, \ldots, \boldsymbol{c}_i, \ldots, \boldsymbol{c}_{i+w}, \boldsymbol{n}^c]$

Candidate-aware Attention Network

- Learn unified candidate-aware user representation from important clicks



- Aggregate clicks relevant to candidate news
- $a_i = \Phi(\mathbf{m}_i, \mathbf{n}^c)$, where Φ is MLP
- $\boldsymbol{u} = \sum_{i=1}^{N} \alpha_i \boldsymbol{m}_i$, $\alpha_i = \frac{\exp(a_i)}{\sum_{j=1}^{N} \exp(a_j)}$

Experimental Datasets

- MIND:
  - A public news recommendation dataset based on Microsoft News
  - Constructed by user logs from 2019.10.19 to 2019.11.15 (6 weeks)
- NewsApp:
  - A news recommendation dataset based on user logs in a Microsoft news App
  - Constructed by user logs from 2020.01.23 to 2020.04.23 (13 weeks)

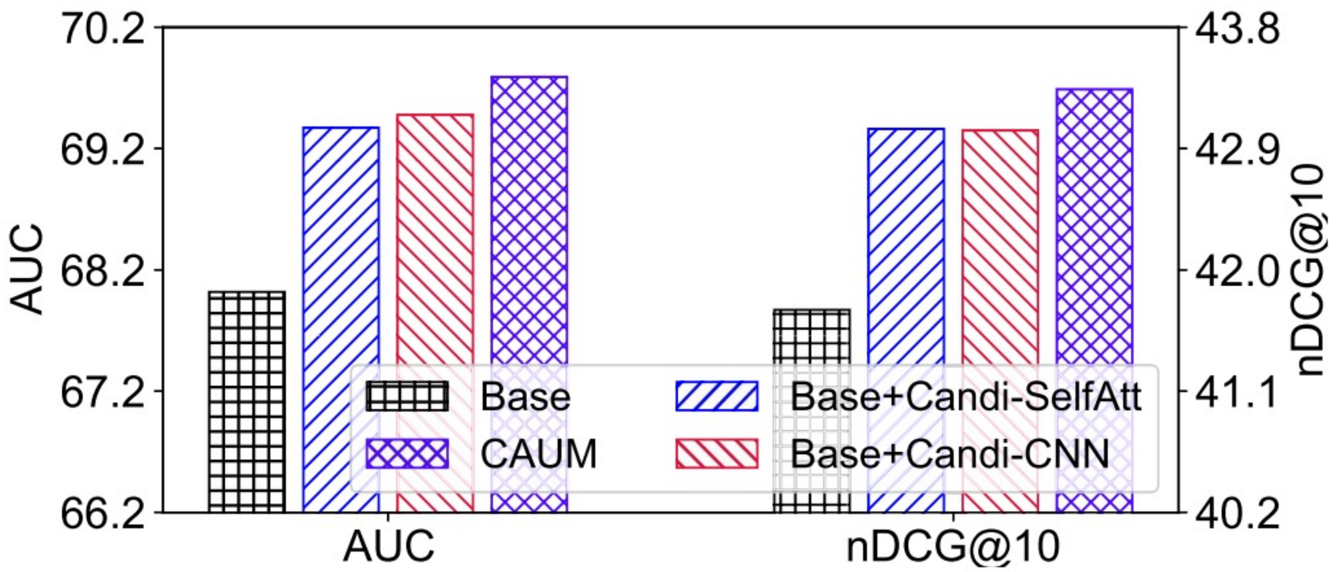
	MIND	*NewsApp*
# News	161,013	1,126,508
# Users	1,000,000	50,605
# Topic Categories	18	28
Avg. # clicks of a user	24.2	19.4
Avg. # words in news title	11.78	11.90
Avg. # entities in news title	2.86	0.99

Performance Evaluation

	MIND				*NewsApp*			
	AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
GRU [10]	65.69±0.15	31.47±0.06	33.96±0.07	39.70±0.07	63.23±0.37	27.83±0.26	31.84±0.31	37.41±0.34
NAML [20]	66.49±0.19	32.38±0.13	35.17±0.15	40.84±0.14	64.52±0.35	29.02±0.20	33.35±0.30	38.90±0.33
NPA [21]	66.56±0.18	32.42±0.10	35.20±0.11	40.87±0.13	64.39±0.14	28.93±0.10	33.31±0.11	38.83±0.11
NRMS [24]	68.04±0.20	33.31±0.07	36.23±0.15	41.92±0.12	65.36±0.28	29.47±0.21	33.96±0.27	39.49±0.19
LSTUR [1]	68.36±0.22	33.30±0.11	36.30±0.16	42.00±0.14	65.18±0.23	29.28±0.21	33.71±0.23	39.28±0.22
KRED [8]	67.73±0.13	32.87±0.11	35.81±0.13	41.43±0.15	65.45±0.14	29.56±0.09	34.11±0.11	39.65±0.12
DKN [19]	66.32±0.18	32.13±0.14	34.86±0.13	40.47±0.18	62.86±0.37	28.00±0.23	32.12±0.29	37.68±0.28
HiFi-Ark [9]	67.93±0.25	32.87±0.07	35.77±0.08	41.47±0.10	64.91±0.15	29.10±0.12	33.52±0.18	38.98±0.14
FIM [18]	67.84±0.12	33.26±0.06	36.18±0.10	41.86±0.11	65.39±0.10	29.63±0.11	34.14±0.12	39.60±0.10
GNewsRec [3]	68.36±0.22	33.41±0.10	36.36±0.13	42.01±0.14	65.31±0.22	29.40±0.14	33.92±0.16	39.48±0.16
CAUM	**70.04**±0.08	**34.71**±0.08	**37.89**±0.07	**43.57**±0.07	**66.44**±0.07	**30.07**±0.10	**34.69**±0.12	**40.23**±0.10

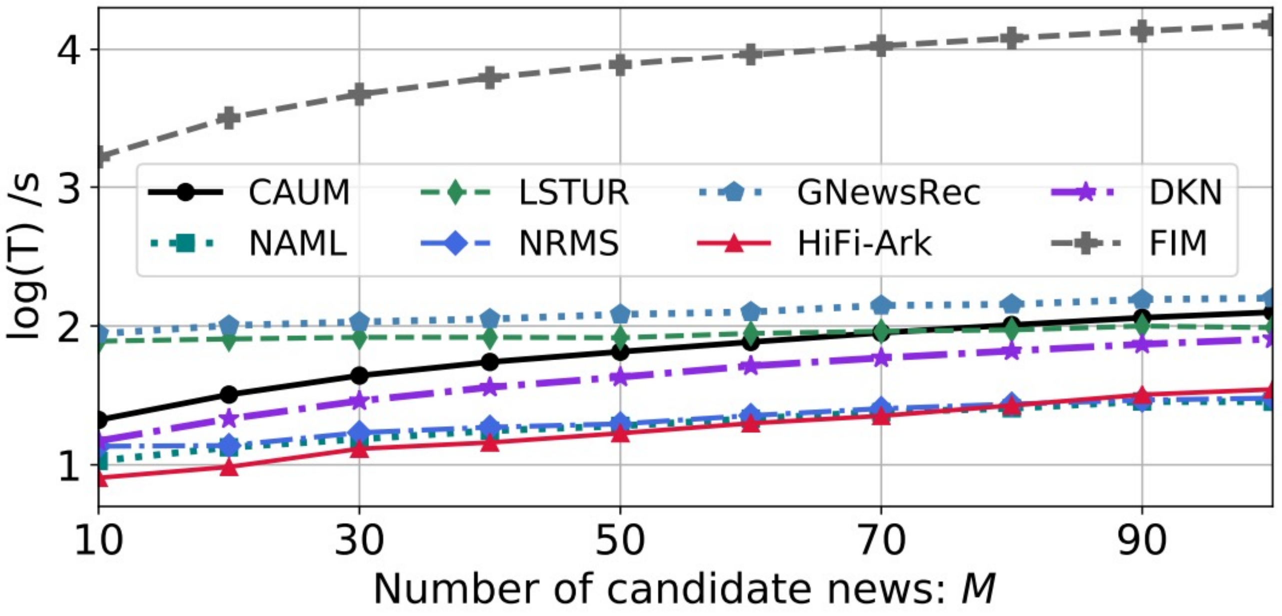
CAUM significantly outperforms baseline user modeling methods at level $p \leq 0.01$.

Ablation Study



Effectiveness of different modules in CAUM.

Analysis on Model Efficiency



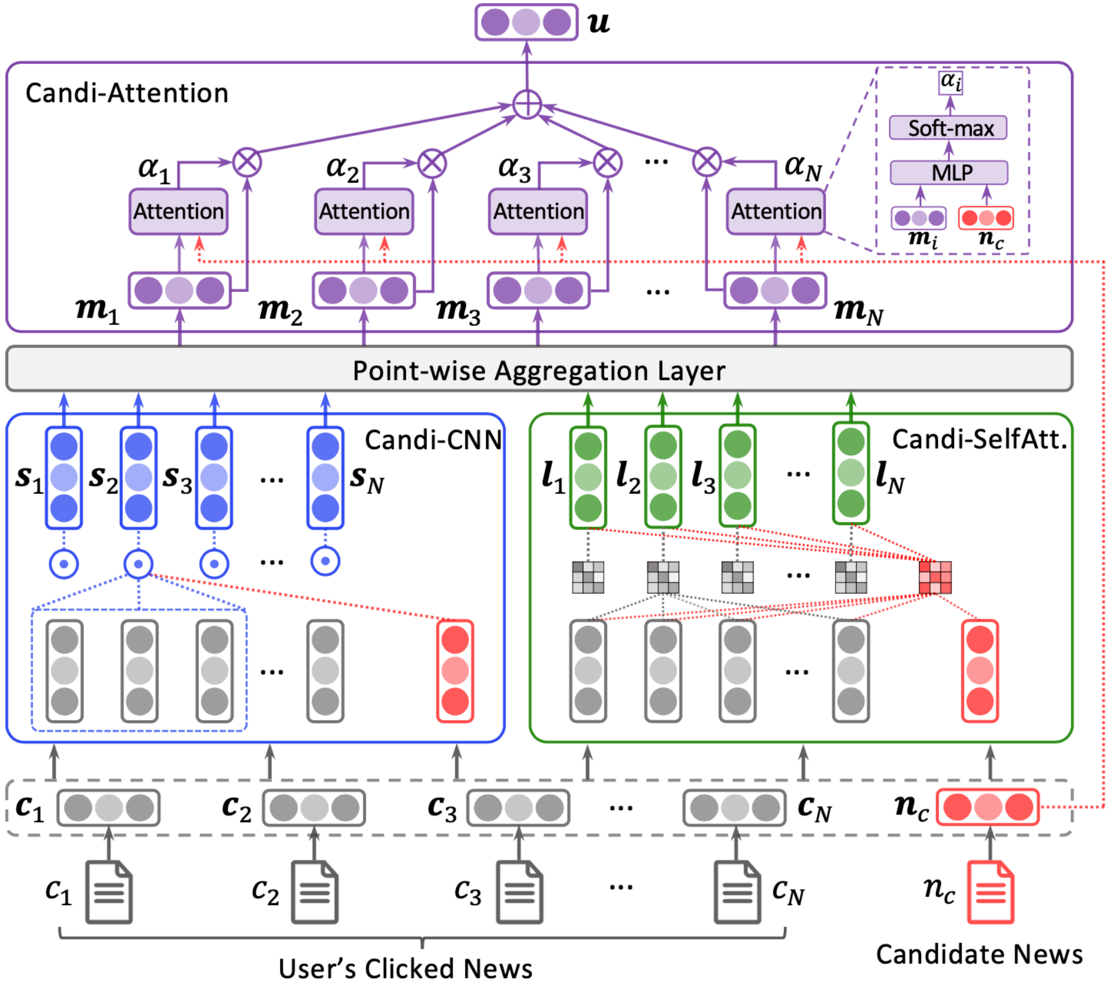
Normalized inference time of different methods.

Method	Time Complexity
NAML	$\mathcal{O}(Md + Nd^2)$
GRU	$\mathcal{O}(Md + Nd^2)$
LSTUR	$\mathcal{O}(Md + Nd^2)$
NRMS	$\mathcal{O}(3Nd^2 + N^2d + Md)$
CAUM	$\mathcal{O}((3N + M)d^2 + (N^2 + MN)d)$

Method time complexity of calculating matching scores of M candidate news.

Conclusion

- A candidate-aware user modeling framework for news recommendation.



Tao Qi

taoqi.qt@gmail.com