News Recommendation with Candidate-aware User Modeling (SIGIR_2022)

Tao Qi
Department of Electronic Engineering,
Tsinghua University
taoqi.qt@gmail.com

Fangzhao Wu
Microsoft Research Asia
wufangzhao@gmail.com

Chuhan Wu
Department of Electronic Engineering,
Tsinghua University
wuchuhan15@gmail.com

Yongfeng Huang
Department of Electronic Engineering,
Tsinghua University
yfhuang@tsinghua.edu.cn

2022. 7. 17 • ChongQing

gesis Leibniz-Institut für Sozialwissenschaften

Forschungszentrum · Research Center L3S

Reported by Lele Duan

Code: https://github.com/taoqi98/CAUM

1.Background

2.Method

3.Experiments

Background

- News recommendation aims to match news with personalized user interest.
- Existing methods model user interests and candidate news content in a candidate-agnostic way.
- Be difficult to accurately match candidate news with a specific user interest if candidate news is not considered in user modeling.

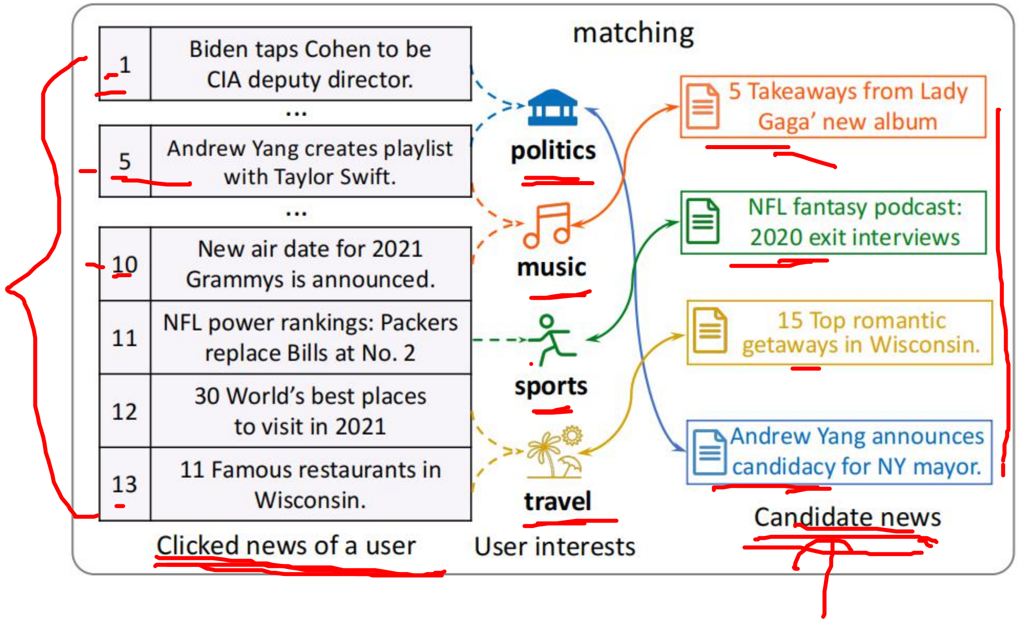

Figure 1: The matching between candidate news and user interest inferred from historical clicked news.

➢ Over view

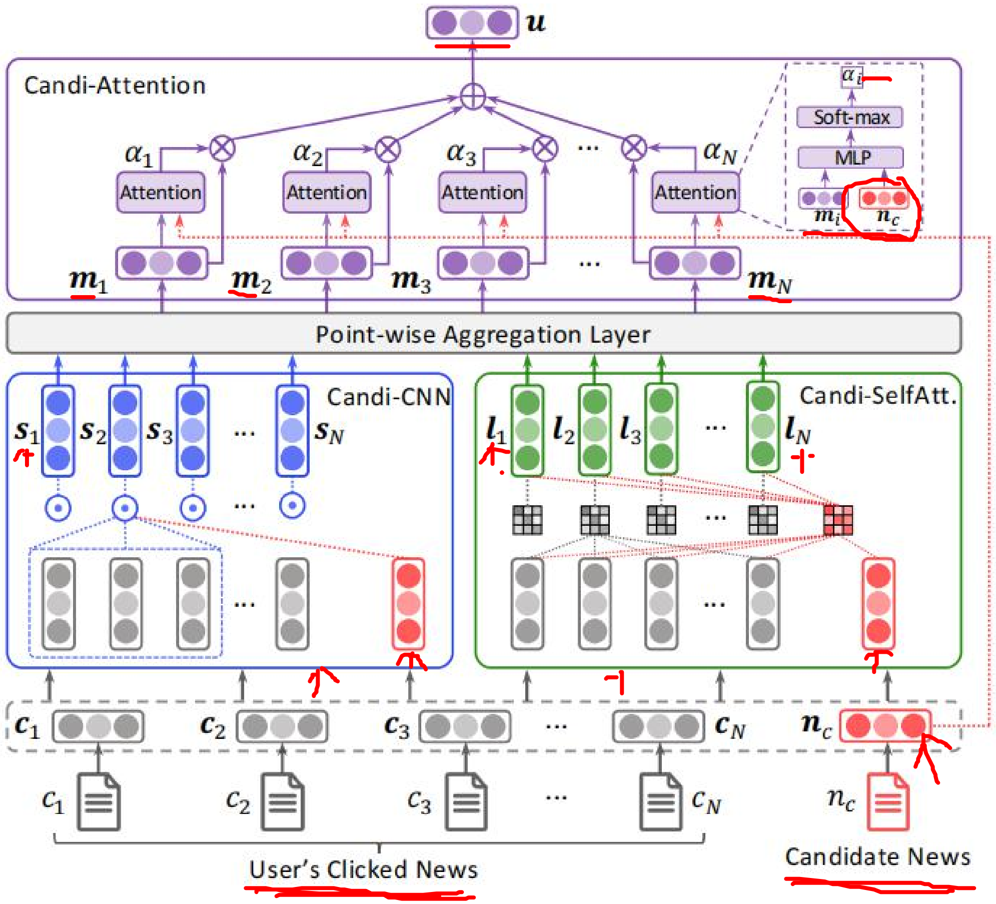

Figure 2: Framework of our *CAUM* method.

Candidate-aware User Modeling

➢ Candi-SelfAt : Modeling global user interests.

$$\hat{r}^k_{i,j} = \mathbf{q}_i^T \mathbf{W}^r_k \mathbf{c}_j, \quad \mathbf{q}_i^T = \mathbf{Q}_u \mathbf{c}_i, \tag{1}$$

where $\hat{r}^k_{i,j}$ denotes the attention score generated by the k-th attention head to model relatedness between the i-th and j-th click, $\mathbf{Q}_u$ is the projection matrix, and $\mathbf{W}^r_k$ is parameters of the k-th attention head. Note that $\{\hat{r}^k_{i,j}\}_{j=1}^N$ model the relatedness between the i-th clicks and other user's clicks. We further adaptively select informative long-range relatedness for modeling user interest in candidate n_c based on their relevance with candidate news:

$$r^k_{i,j} = \hat{r}^k_{i,j} + \mathbf{q}_c^T \mathbf{W}^r_k \mathbf{c}_j, \quad \mathbf{q}_c^T = \mathbf{Q}_c \mathbf{n}_c, \tag{2}$$

where $r^k_{i,j}$ is the candidate-aware attention score, and $\mathbf{Q}_c$ is a projection matrix. Then we learn the representation $\mathbf{l}^k_i$ generated by the k-th head for the i-th click based on attention weights $\{\gamma^k_j\}_{j=1}^N$:

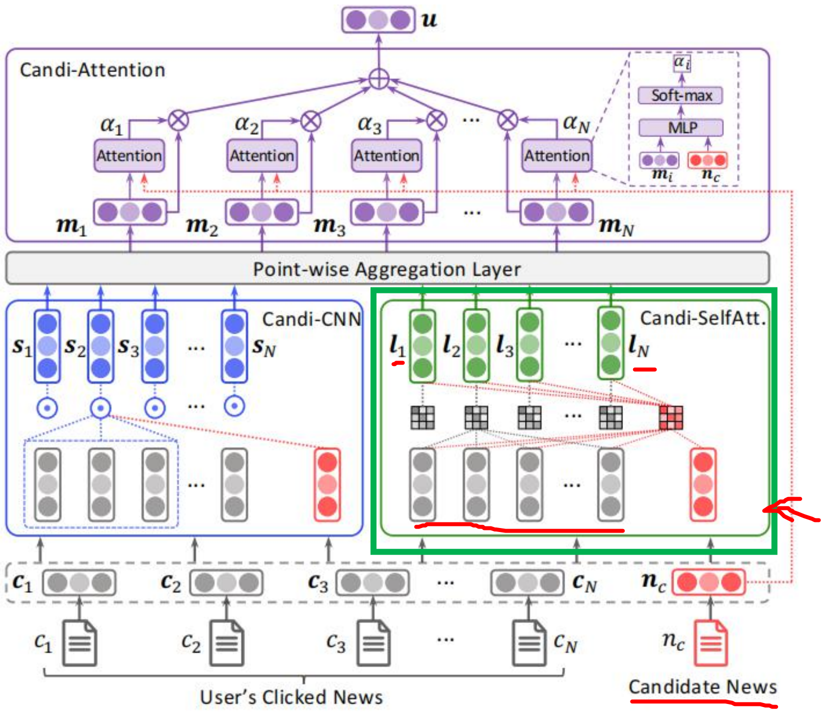

Figure 2: Framework of our *CAUM* method.

$$\mathbf{l}^k_i = \mathbf{W}^k_o \sum_{j=1}^N \gamma^k_j \mathbf{c}_j, \quad \gamma^k_j = \frac{\exp(r^k_{i,j})}{\sum_{p=1}^N \exp(r^k_{i,p})}, \tag{3}$$

where $\mathbf{W}^k_o$ is the projection matrix of the k-th attention head. Finally, we learn the global contextual representation $\mathbf{l}_i$ for i-th click by contacting $\{\mathbf{l}^k_i\}_{k=1}^K$, where K is the number of attention heads.

Candidate-aware User Modeling

➢ Candi-CNN : Modeling short-term user interests.

$$\mathbf{s}_i = \mathbf{W}_c[\mathbf{c}_{i-h}; ...; \mathbf{c}_i; ...; \mathbf{c}_{i+h}; \mathbf{n}_c],$$

where $\mathbf{s}_i$ represents local contextual representation of the i-th click, $2h+1$ is the window size of the CNN network, and $\mathbf{W}_c$ represents parameters of filters in the Candi-CNN network.

Local contextual representations:

$$[\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N]$$

$\mathbf{m}_i = \mathbf{P}_m[\mathbf{s}_i; \mathbf{l}_i]$, where $\mathbf{P}_m$ is the projection matrix.

➢ Candi-Att:

$$\mathbf{u} = \sum_{i=1}^{N} \alpha_i \mathbf{m}_i, \ \alpha_i = \frac{\exp(\Phi(\mathbf{m}_i, \mathbf{n}_c))}{\sum_{j=1}^{N} \exp(\Phi(\mathbf{m}_j, \mathbf{n}_c))}, \qquad (4)$$

where α_i is the weight of the i-th click and Φ is an MLP network. In this way, user interests relevant to the candidate news can be effectively encoded into $\mathbf{u}$ to improve the accuracy of interest matching.

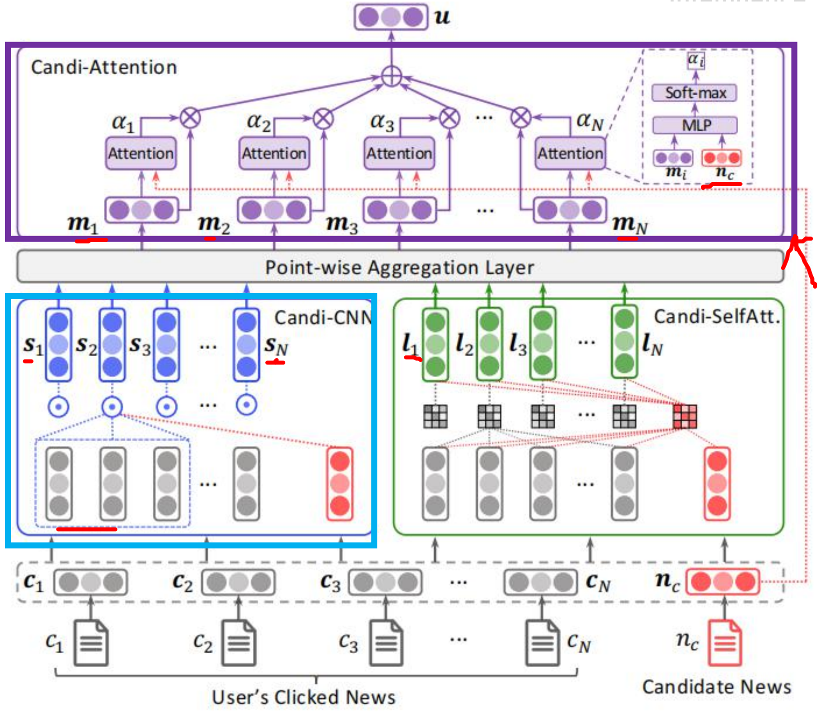

Figure 2: Framework of our *CAUM* method.

News Modeling:

$$\mathbf{n} = \mathbf{n}^t + \mathbf{n}^e + \mathbf{n}^v.$$

Interest Matching and Model Learning:

$$\hat{y} = \mathbf{n}_c^T \cdot \mathbf{u}$$

$$\mathcal{L} = -\frac{1}{H}\sum_{i=1}^{H} \log \phi(\hat{y}_i^p - \hat{y}_i^n)$$

Experiment

Table 1: Performance comparisons. The improvement of *CAUM* over baseline methods is significant at level $p \leq 0.001$.

	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

Table 2: Method time complexity (multiplication operation) of calculating matching scores of M candidate news. News and user representation are d-dimensional.

NAML	$O(Md + Nd^2)$	GRU	$O(Md + Nd^2)$
LSTUR	$O(Md + Nd^2)$	NRMS	$O(3Nd^2 + N^2d + Md)$
CAUM	$O((3N + M)d^2 + (N^2 + MN)d)$		

Experiment

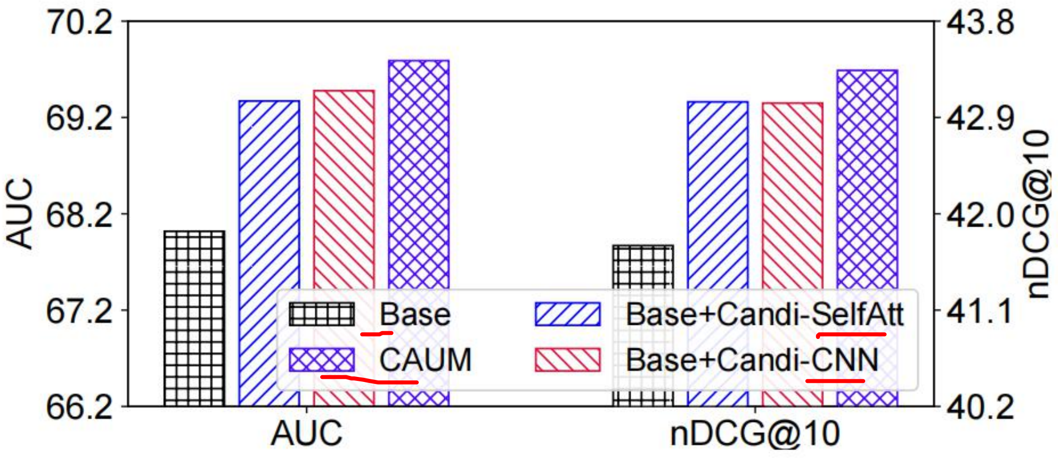

Figure 3: Ablation study of *CAUM*.

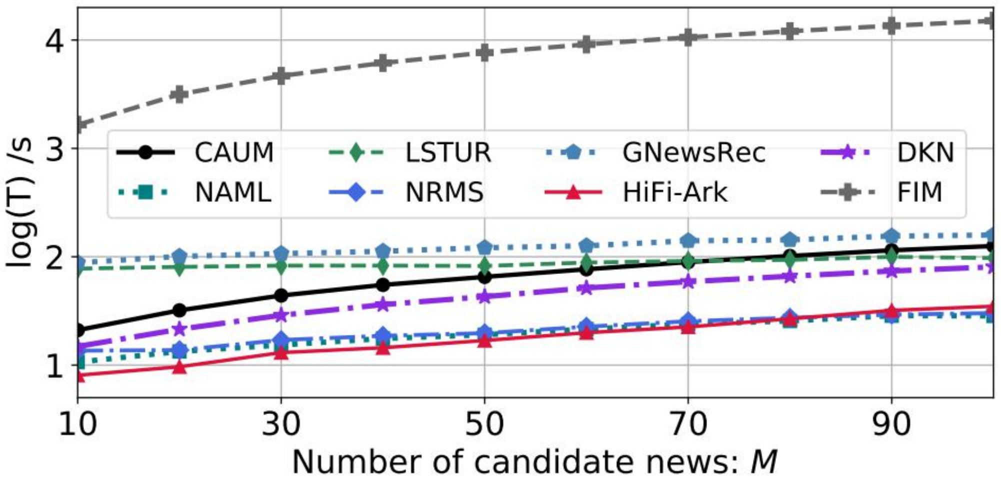

Figure 4: Normalized inference time of different methods.

Thank you!