FUM: Fine-grained and Fast User Modeling for News Recommendation

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

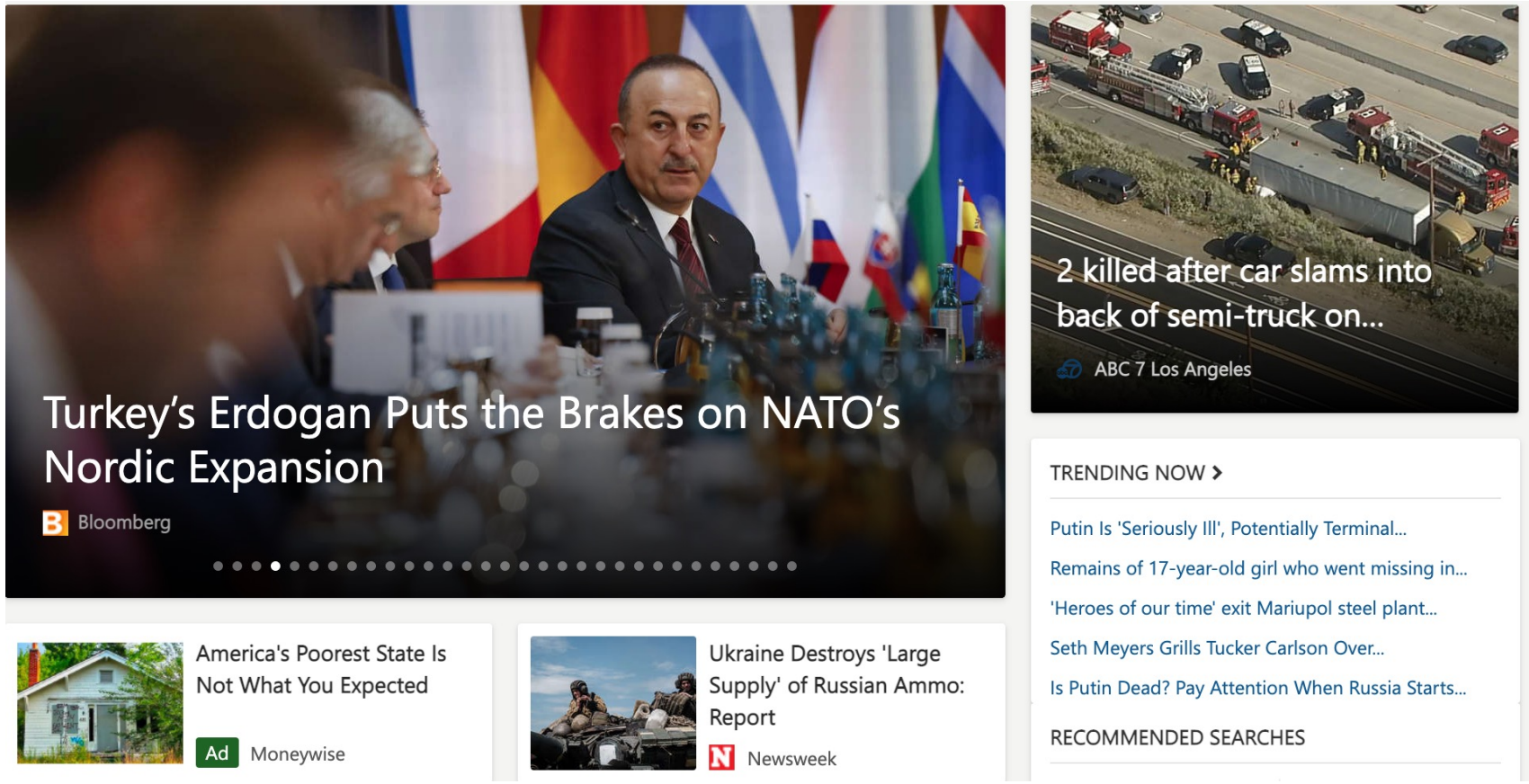
[1]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

[2]Microsoft Research Asia, Beijing 100080, China

taoqi.qt@gmail.com

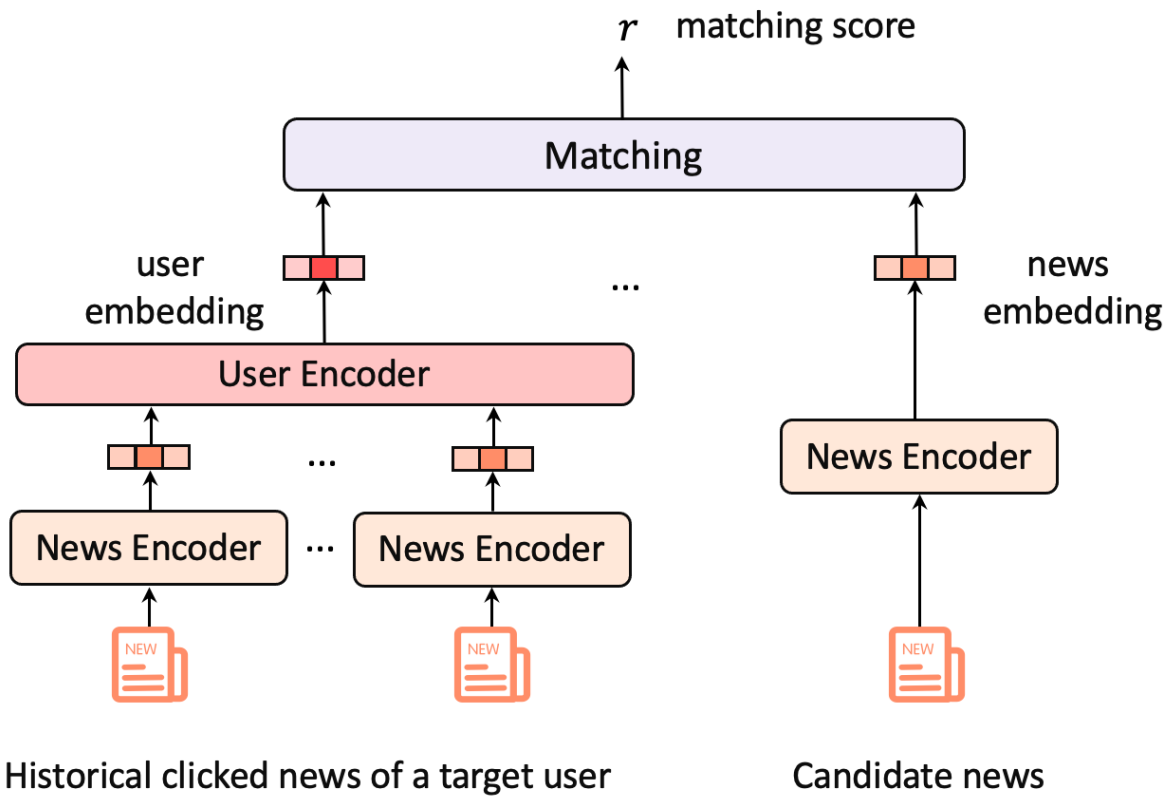
Personalized News Recommendation

- News recommendation can effectively alleviate information overload
- The key of news recommendation is to accurately model user interest



Mainstream User Modeling Framework

- Existing methods usually independently encode user's clicked news into news embeddings and then aggregate them into user embedding



Mainstream user modeling framework for news recommendation.

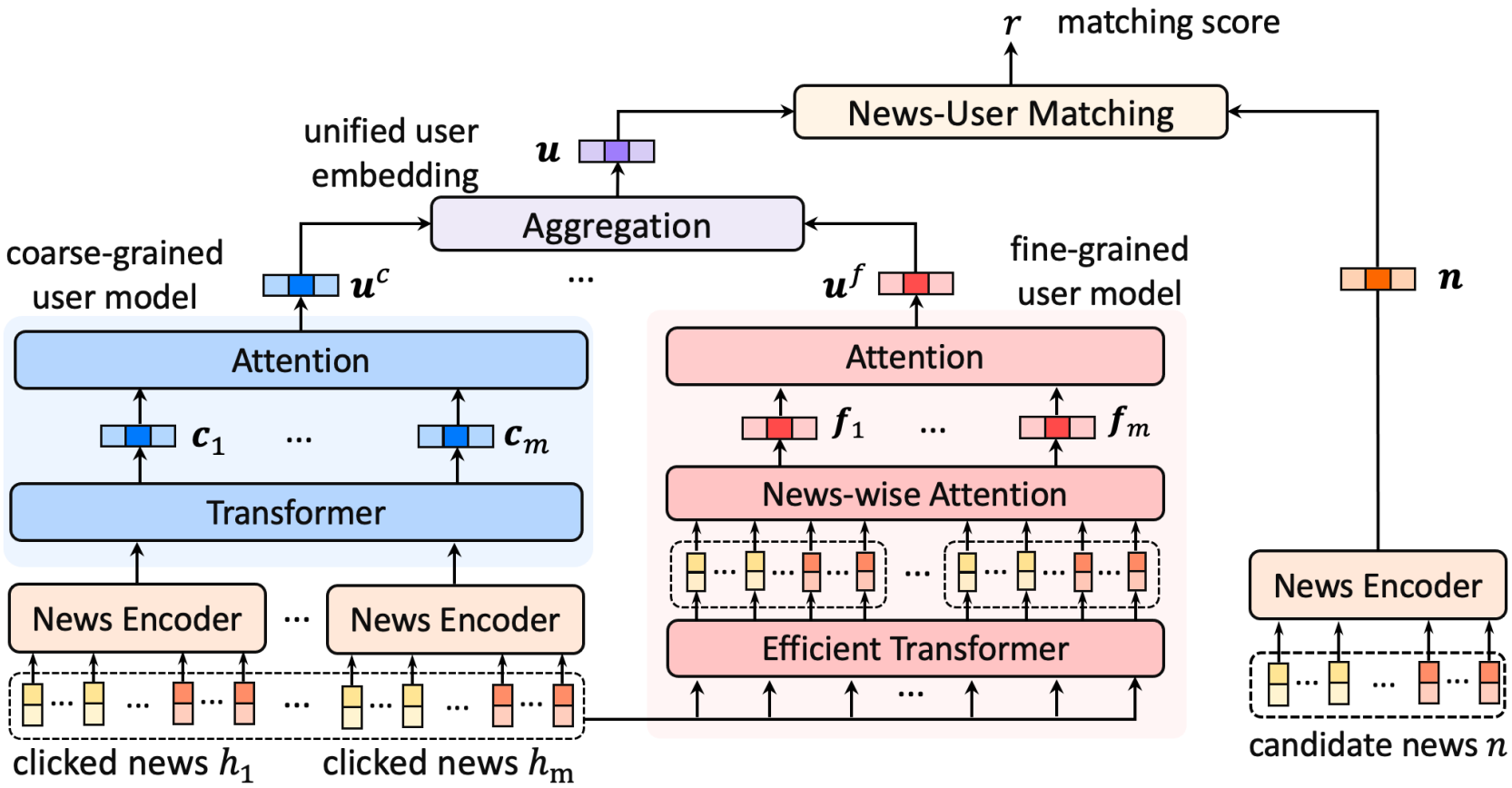
Fine-grained User Modeling For News Rec.

- The word-level interactions across clicked news is important for user modeling
- Most of the existing methods neglect word-level behavior interactions

	Texts of user's clicked news
1	The challenge in the new story for Iron Man.
2	The upcoming movies of Netflix in 2022.
3	The success of Marvel's Avengers.
4	Adele says if 30 doesn't come now it never will.
5	The most popular songs on YouTube in this week.
6	Vinyl and CD sales both went up in 2021, data says.

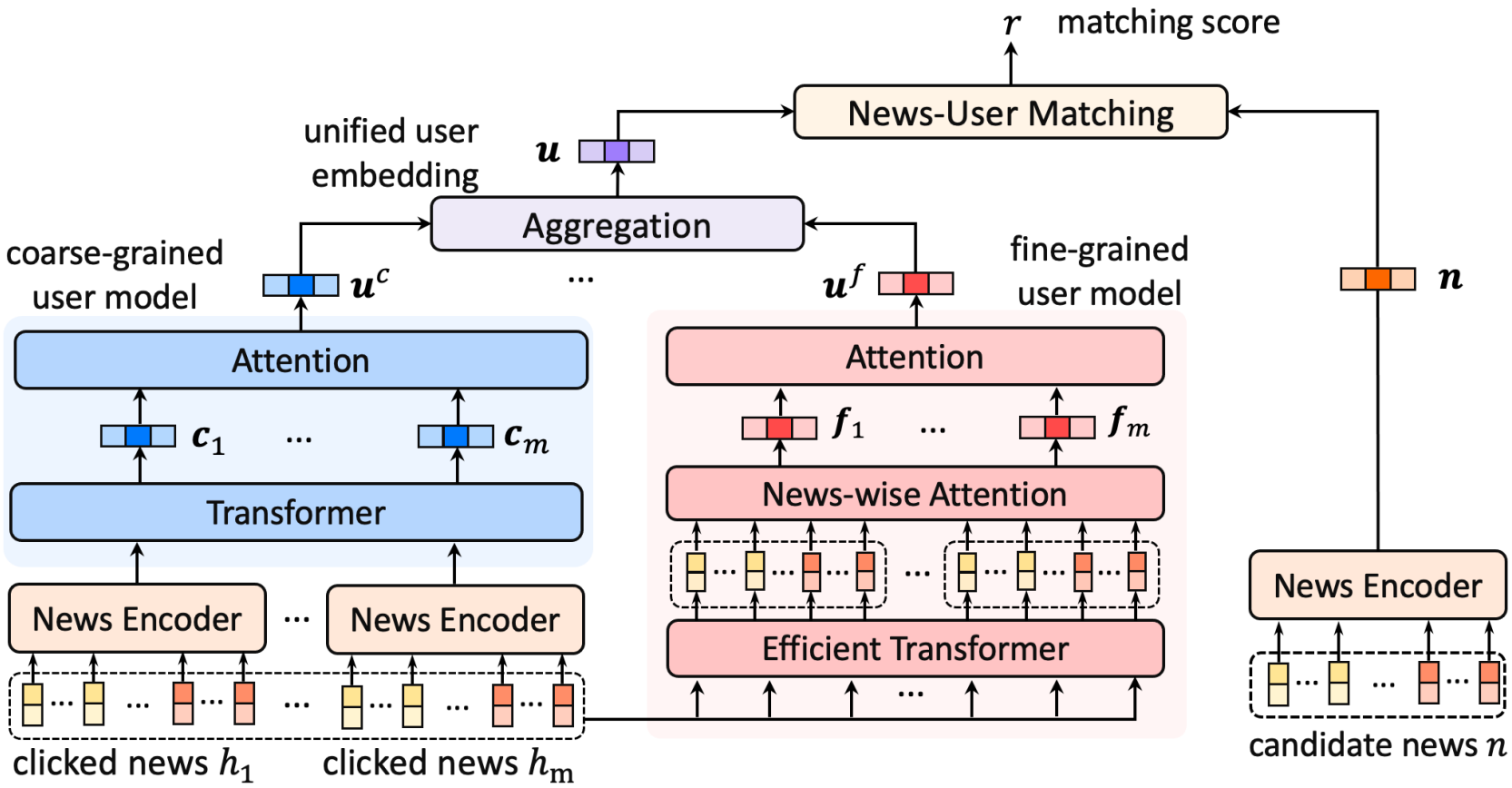
FUM: Fine-grained and Fast User Modeling

- Transform user modeling into a document modeling task



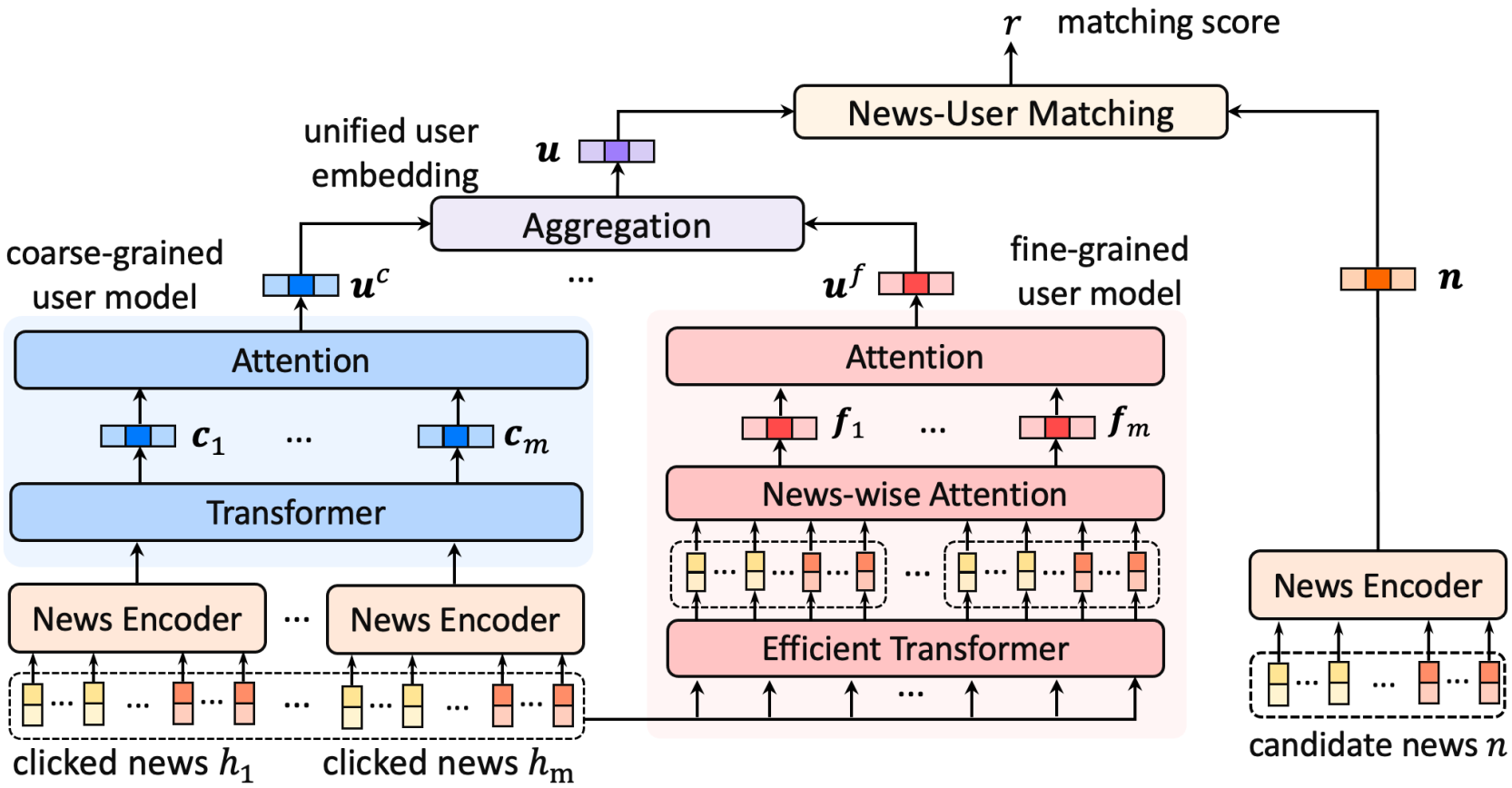
FUM: Fine-grained and Fast User Modeling

- **Fine-grained user model**: capture user interest from word-level behavior interactions via efficient transformers



FUM: Fine-grained and Fast User Modeling

- **Coarse-grained user model**: summarize user interest from news-level behavior interactions via a hierarchical pooling network



Experimental Datasets and Settings

- The MIND dataset
  - A public news recommendation dataset based on Microsoft News
  - Constructed by user logs from 2019.10.19 to 2019.11.15 (6 weeks)
- The Feeds dataset
  - A news recommendation dataset based on user logs in a Microsoft news feeds
  - Constructed by user logs from 2020.01.23 to 2020.04.23 (13 weeks)
- Hyper-parameter settings
  - Using the Fastformer network as the efficient transformer
  - 400-dimensional news and user embeddings
  - Using Adam with e-4 learning rate for model training

Performance Evaluation

	MIND				*Feeds*			
	AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
GRU	65.47±0.18	31.15±0.22	33.64±0.24	39.34±0.24	62.95±0.13	27.57±0.08	31.55±0.12	37.18±0.11
DKN	67.19±0.13	32.97±0.19	35.87±0.22	41.53±0.17	64.02±0.25	28.65±0.13	32.97±0.17	38.54±0.17
NPA	67.42±0.15	32.97±0.18	35.90±0.23	41.54±0.20	64.83±0.47	29.21±0.36	33.64±0.47	39.18±0.48
KRED	67.77±0.15	33.39±0.15	36.34±0.17	42.04±0.15	64.92±0.14	29.27±0.08	33.71±0.13	39.25±0.12
GNewsRec	68.38±0.09	33.46±0.22	36.44±0.23	42.15±0.20	65.02±0.11	29.28±0.10	33.74±0.13	39.28±0.13
NAML	68.16±0.11	33.31±0.07	36.26±0.10	41.94±0.08	65.31±0.12	29.47±0.07	33.99±0.09	39.57±0.12
NRMS	68.33±0.27	33.55±0.27	36.53±0.32	42.18±0.30	65.21±0.12	29.39±0.05	33.87±0.06	39.46±0.08
LSTUR	68.53±0.10	33.58±0.15	36.54±0.18	42.23±0.17	65.31±0.20	29.54±0.15	34.08±0.19	39.63±0.19
FIM	68.15±0.33	33.36±0.27	36.38±0.30	42.02±0.31	65.47±0.12	29.62±0.07	34.19±0.09	39.72±0.09
FUM	**70.01**±0.10	**34.51**±0.13	**37.68**±0.14	**43.38**±0.13	**66.93**±0.19	**30.49**±0.16	**35.31**±0.21	**40.87**±0.18

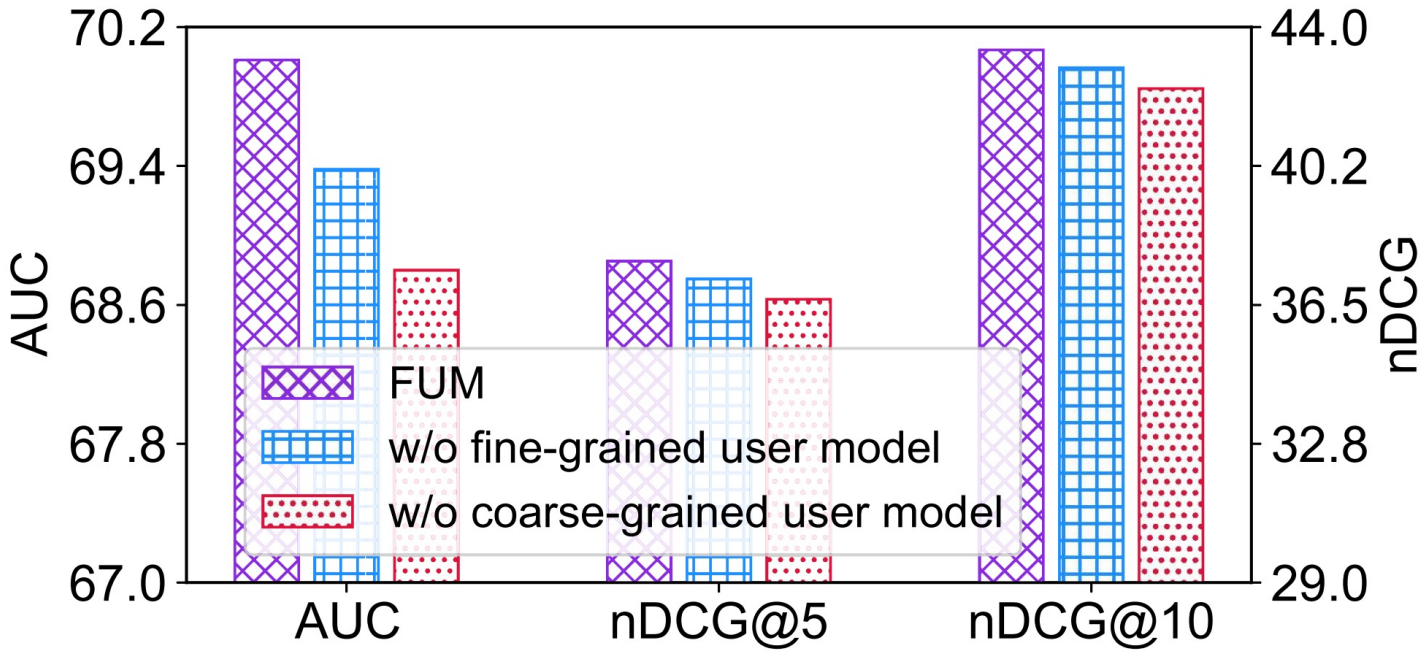
FUM significantly outperforms baseline user modeling methods at level $p \leq 0.01$.

Efficiency Comparison

	GRU	DKN	NAML	NPA	KRED	GNewsRec	LSTUR	NRMS	FIM	FUM
Training Time	11.46s	8.19s	7.98s	8.10s	10.40s	10.72s	11.53s	11.39s	15.85s	13.21s
Inference Time	2.41s	44.90s	1.23s	1.15s	1.24s	86.90s	2.43s	2.16s	350.38s	2.75s
Cacheable	✓	✗	✓	✓	✓	✗	✓	✓	✗	✓

FUM achieves comparable or better efficiency than methods that neglects fine-grained behavior interactions.

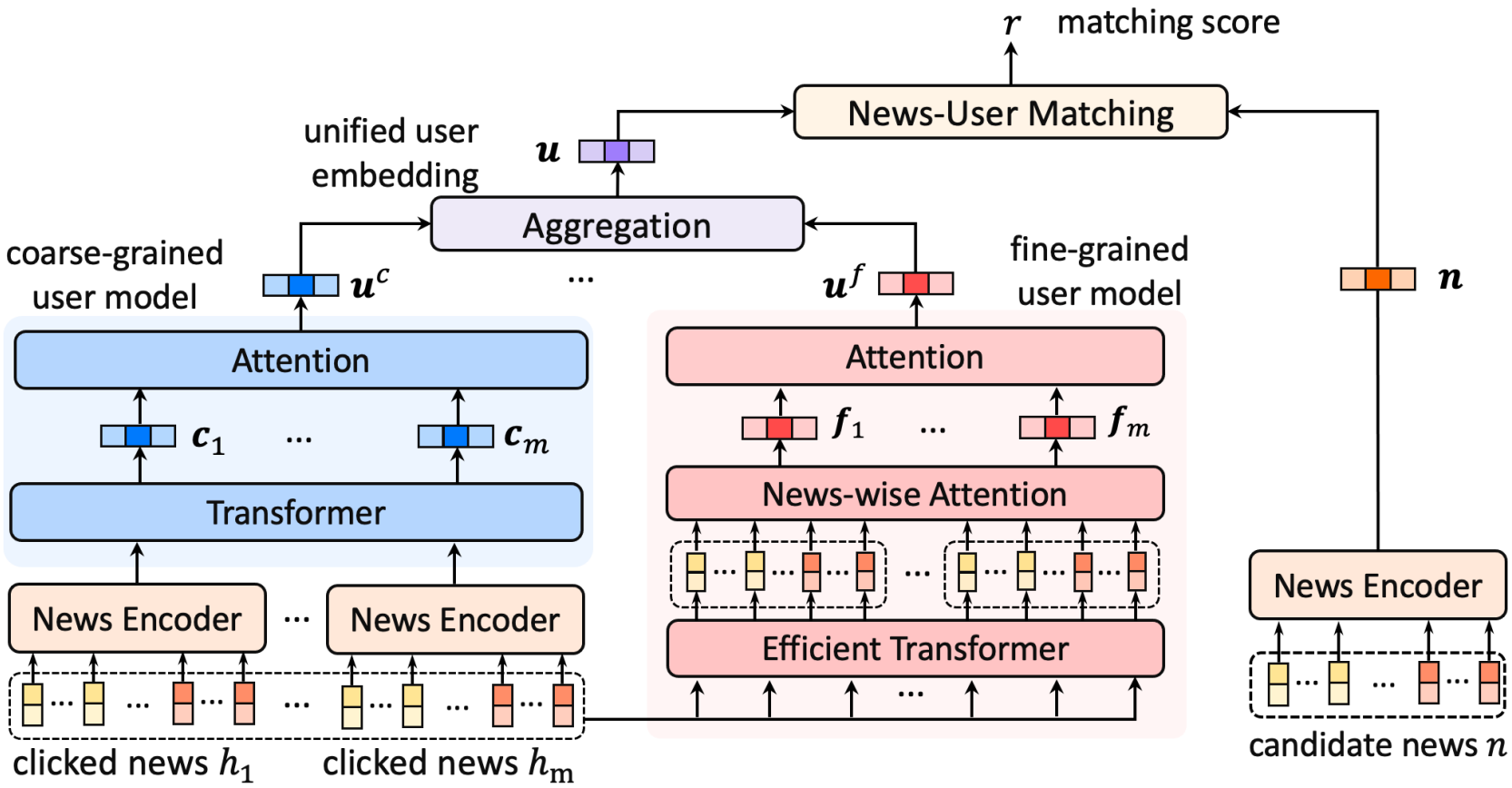
Ablation Study



Both fine- and coarse- grained user models improve the effectiveness of FUM.

Conclusion

- A fine-grained and fast user modeling method for news recommendation



Tao Qi

taoqi.qt@gmail.com