HieRec: Hierarchical User Interest Modeling for Personalized News Recommendation

ADVISOR: JIA-LING KOH
PRESENTER: Xiao-Yuan Hung
SOURCE: ACL'21
DATE: 2022/3/15

Outline

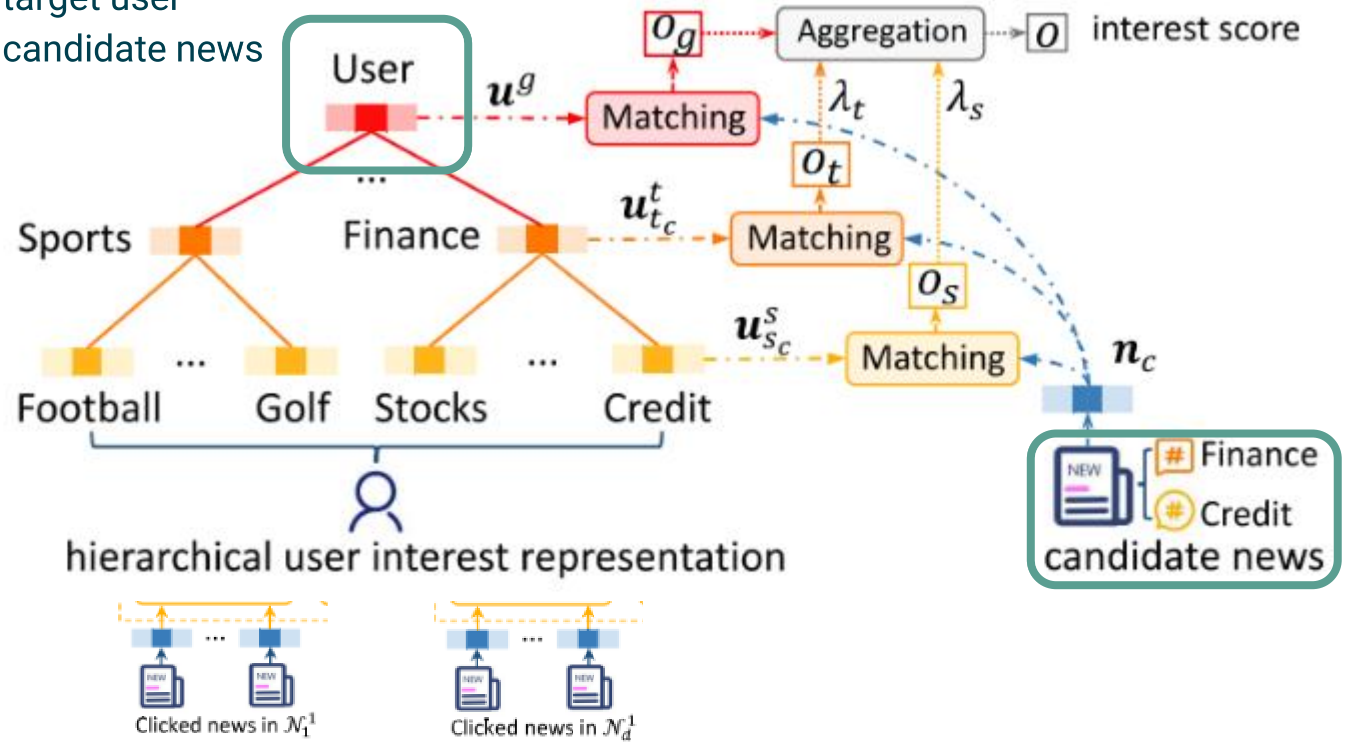
Motivation

- Most existing methods learn a single user embedding to represent the overall user interests.
- However, user interests are usually very diverse and multi-grained, which are difficult to be accurately modeled by a single user embedding.

ID	Click	Topic	Subtopic	Title
1	✓	Movies	Celebrity	Ben Affleck breaks silence after "slip" in sobriety.
2	✓	Sports	Football	Myles Garrett suspended indefinitely by the NFL.
3	✓	Sports	Football	Jaguars veteran cornerback Josh Robinson retires suddenly.
4	✕	Sports	Basketball	Trey Lyles off to a good start with Spurs.
5	✕	Sports	Golf	Can an amateur win again on the PGA Tour?
6	✓	Finance	Stocks	3M is a dog of the Dow -- and it may not get better in 2020.
7	✓	Finance	Taxes	New Trump tax documents show major inconsistencies.
8	✓	Health	Fitness	This guy lost 30 pounds and gained a rock-hard pack.
9	✕	Lifestyle	Food	Candy Corn has tired to the Metro East.

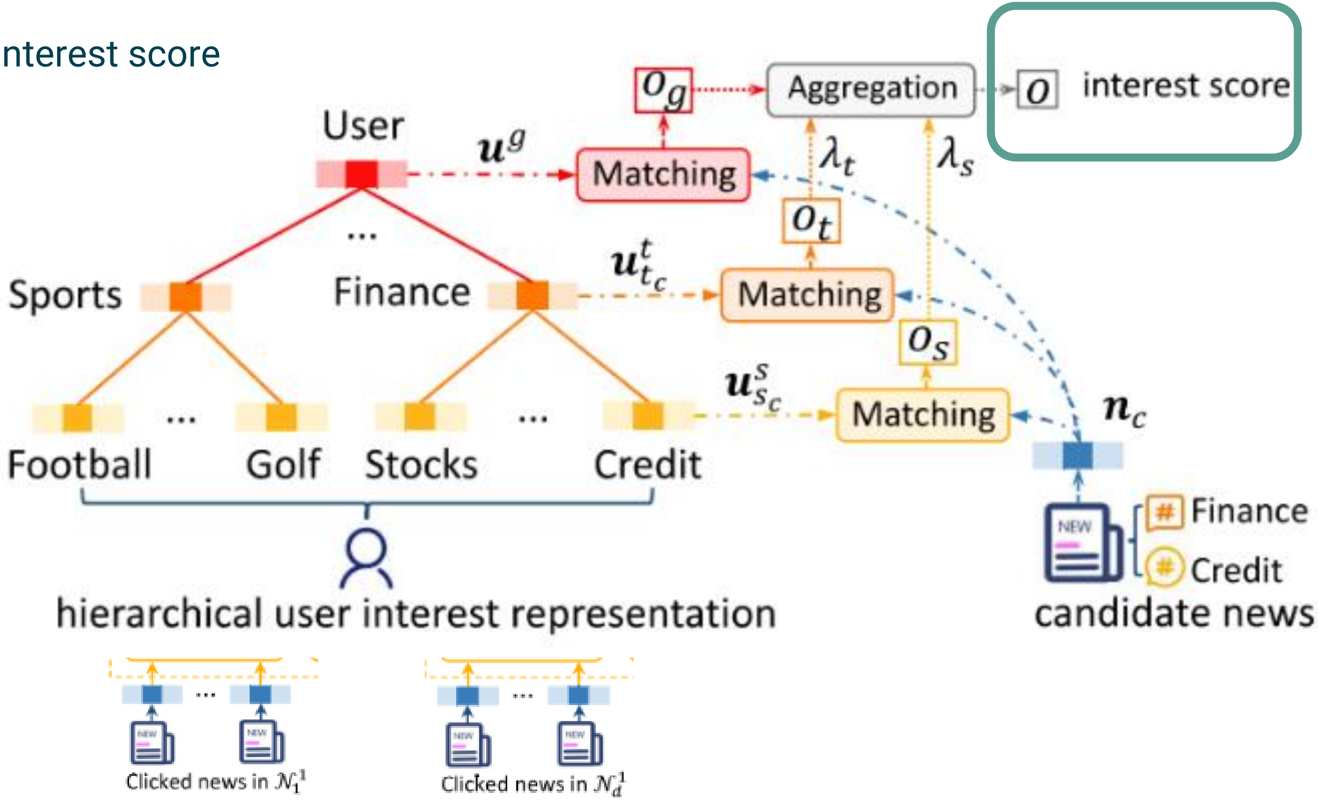
Input

- target user
- candidate news



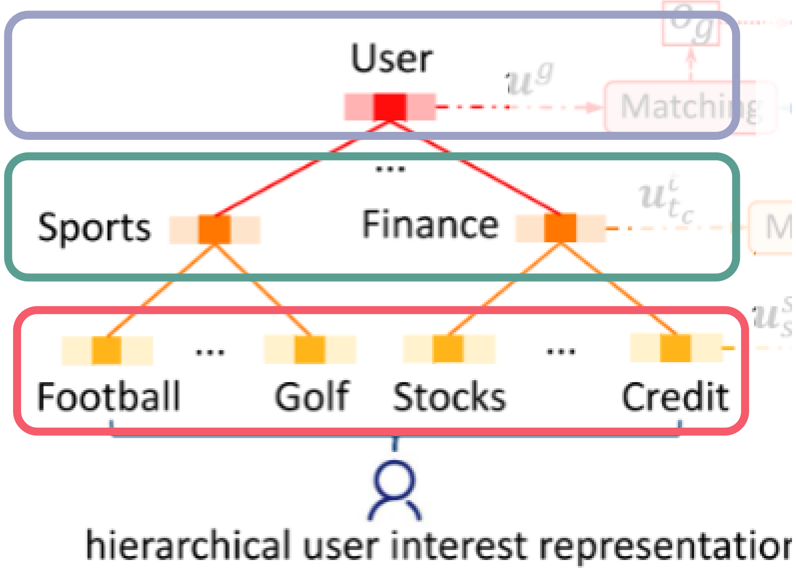
Output

- interest score



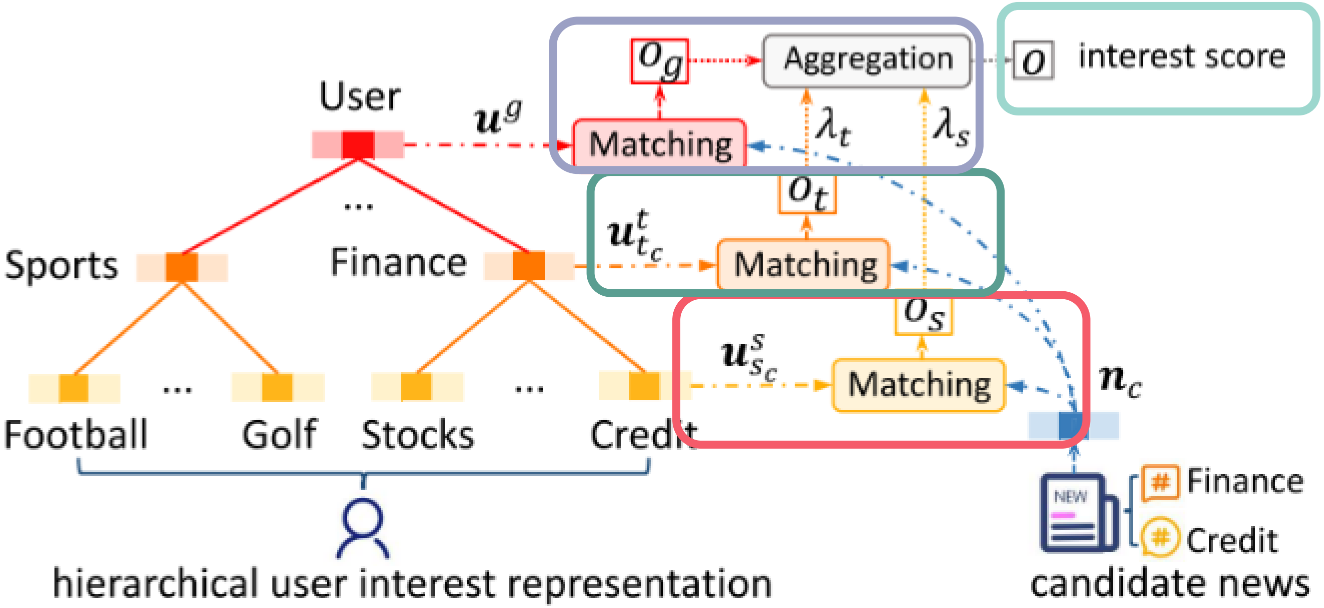
Subtopic-level

ID	Click	Topic	Subtopic	Title
1	✓	Movies	Celebrity	Ben Affleck breaks silence after "slip" in sobriety.
2	✓	Sports	Football	Myles Garrett suspended indefinitely by the NFL.
3	✓	Sports	Football	Jaguars veteran cornerback Josh Robinson retires suddenly.
4	✗	Sports	Basketball	Trey Lyles off to a good start with Spurs.
5	✗	Sports	Golf	Can an amateur win again on the PGA Tour?
6	✓	Finance	Stocks	3M is a dog of the Dow -- and it may not get better in 2020.
7	✓	Finance	Taxes	New Trump tax documents show major inconsistencies.
8	✓	Health	Fitness	This guy lost 30 pounds and gained a rock-hard pack.
9	✗	Lifestyle	Food	Candy Corn has tired to the Metro East.



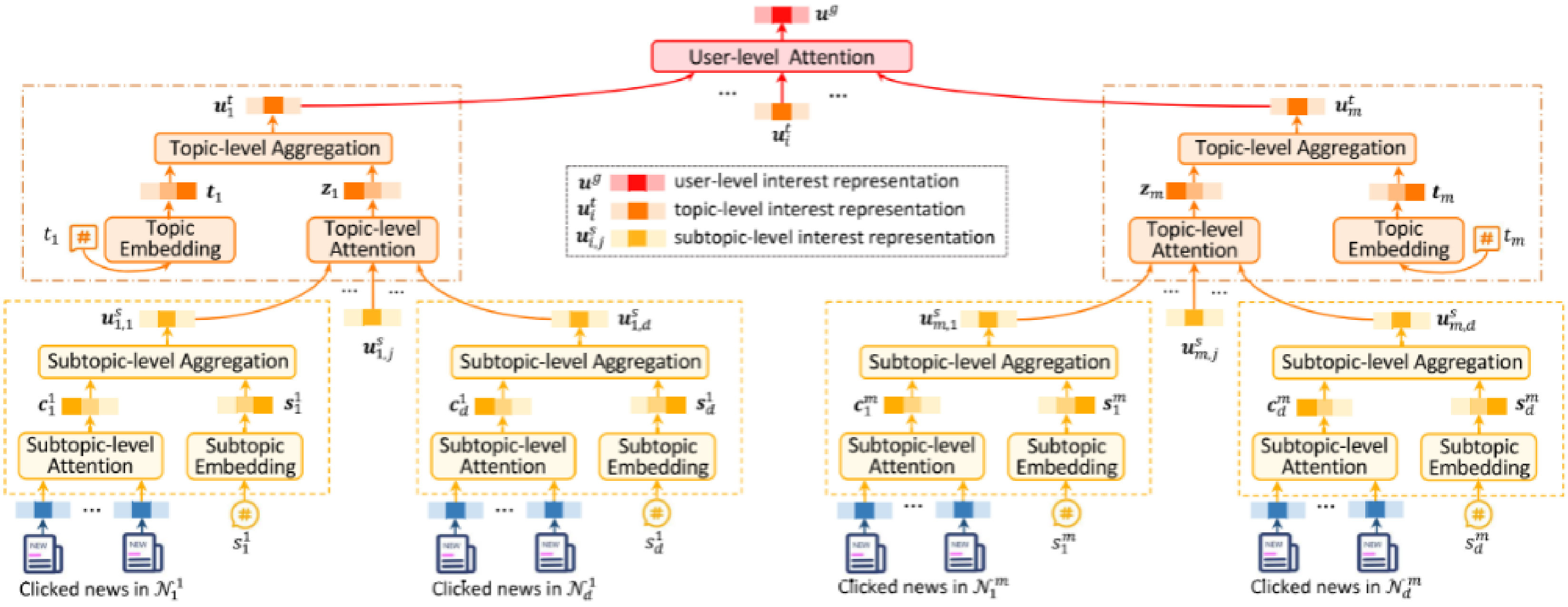
- interest representation to model overall user interests
- coarse-grained user interests in major news topics
- fine-grained user interests in different news subtopics

Interest Matching



Outline

Hierarchical User Interest Modeling



Hierarchical User Interest Modeling

$\mathcal{N}_j^i = \{n_k^{i,j} | k = 1, ..., l\}$

user's clicked news in topic *i* and subtopic *j* are divided into the same click group

clicked topic set $\{t_i | i = 1, ..., m\}$

the user has *M* clicked news.

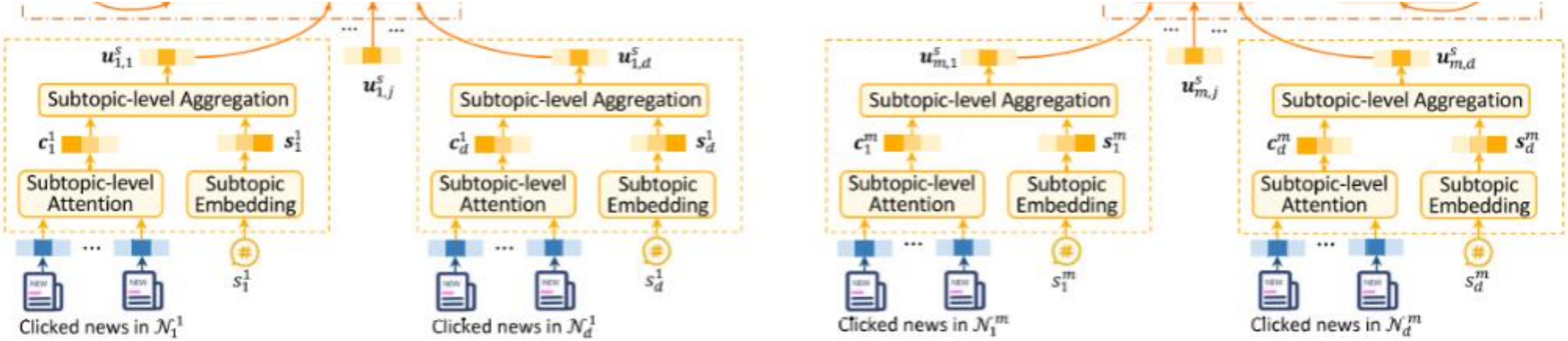
clicked subtopic set $\{s_j^i | j = 1, ..., d\}$

d is the size of the set.

subtopic-level interest representations

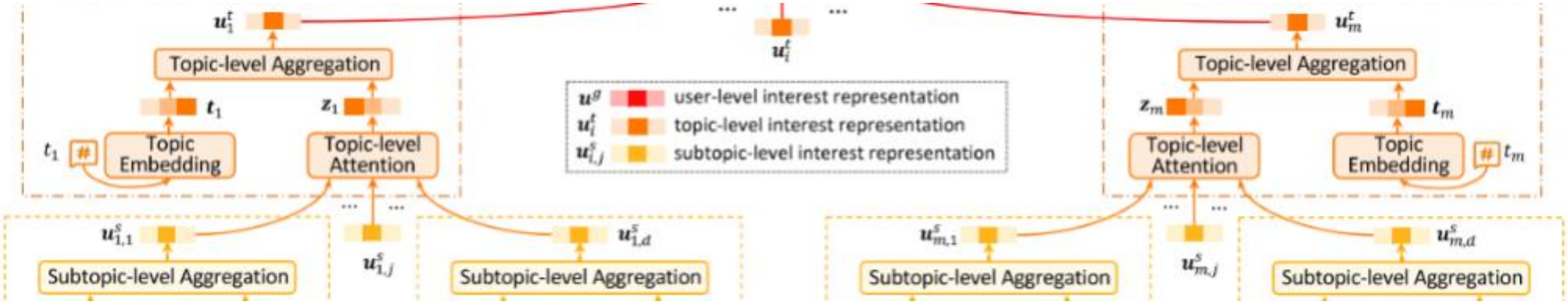
$$\mathbf{c}_j^i = \sum_{k=1}^{l} \gamma_k \mathbf{n}_k^{i,j}, \ \gamma_k = \frac{\exp(\phi_s(\mathbf{n}_k^{i,j}))}{\sum_{p=1}^{l} \exp(\phi_s(\mathbf{n}_p^{i,j}))}, \quad (1)$$

$$\mathbf{u}_{i,j}^s = \mathbf{c}_j^i + \mathbf{s}_j^i.$$



Topic-level interest representations

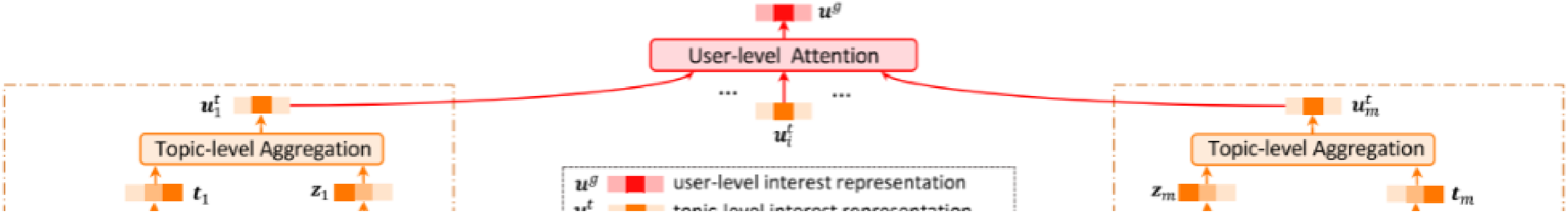
$$\mathbf{u}_i^t = \mathbf{z}_i + \mathbf{t}_i$$



$$\mathbf{z}_i = \sum_{j=1}^{d} \beta_j \mathbf{u}_{i,j}^s, \ \beta_j = \frac{\exp(\phi_t(\mathbf{v}_{i,j}^s))}{\sum_{k=1}^{d} \exp(\phi_t(\mathbf{v}_{i,k}^s))}, \quad (2)$$

$\mathbf{v}_{i,j}^s = [\mathbf{u}_{i,j}^s; \mathbf{r}_j^i]$ $\mathbf{r}_j^i$ is the embedding vector for the number of clicked news on subtopic s_j^i

User-level interest representations



$$\mathbf{u}^g = \sum_{i=1}^{m} \alpha_i \mathbf{u}_i^t, \quad \alpha_i = \frac{\exp(\phi_g(\mathbf{v}_i^t))}{\sum_{j=1}^{m} \exp(\phi_g(\mathbf{v}_j^t))}, \quad (3)$$

$\mathbf{v}_i^t = [\mathbf{u}_i^t; \mathbf{r}_i]$ $\mathbf{r}_i$ is the embedding vector for the number of user's clicked news on topic t_i

News representation learning framework

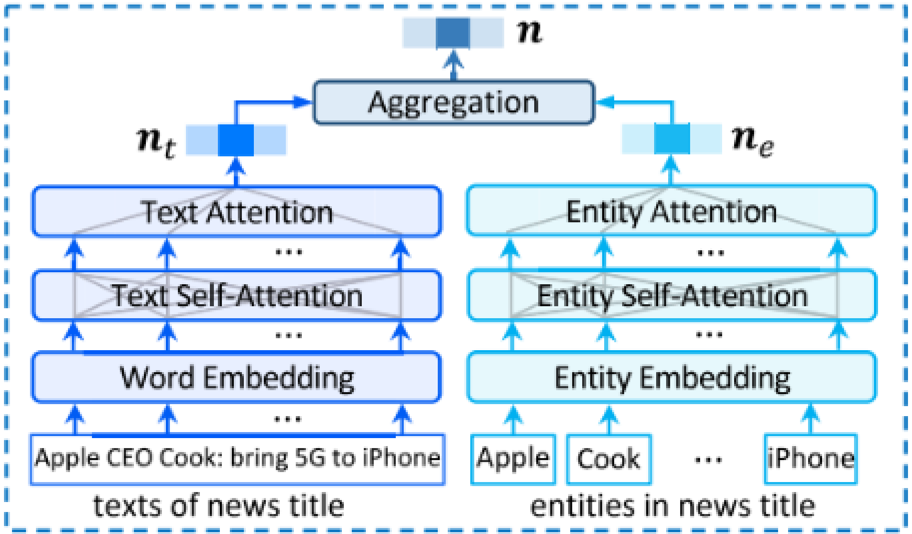


Figure 4: News representation learning framework.

News include

- topice t
- subtopic s
- title
  - text sequence

$$\mathbf{T} = [w_1, w2, ..., w_T]$$

  - entity sequence

$$\mathbf{E} = [e_1, e_2, ..., e_E]$$

$\mathbf{n}_t$:text representation

$\mathbf{n}_e$:entity representation

$$\mathbf{n} = \mathbf{W}_t\mathbf{n}_t + \mathbf{W}_e\mathbf{n}_e$$

where $\mathbf{W}_t$ and $\mathbf{W}_e$ are parameters.

Hierarchical User Interest Matching

- fine-grained user interests is useful for personalized news recommendations
- **baseball** cannot match any fine-grained interests
- coarse-grained user interests and overall user interests can match with it.

ID	Click	Topic	Subtopic	Title
1	✓	Movies	Celebrity	Ben Affleck breaks silence after “slip” in sobriety.
2	✓	Sports	Football	Myles Garrett suspended indefinitely by the NFL.
3	✓	Sports	Football	Jaguars veteran cornerback Josh Robinson retires suddenly.
4	✗	Sports	Basketball	Trey Lyles off to a good start with Spurs.
5	✗	Sports	Golf	Can an amateur win again on the PGA Tour?
6	✓	Finance	Stocks	3M is a dog of the Dow -- and it may not get better in 2020.
7	✓	Finance	Taxes	New Trump tax documents show major inconsistencies.
8	✓	Health	Fitness	This guy lost 30 pounds and gained a rock-hard pack.
9	✗	Lifestyle	Food	Candy Corn has tired to the Metro East.

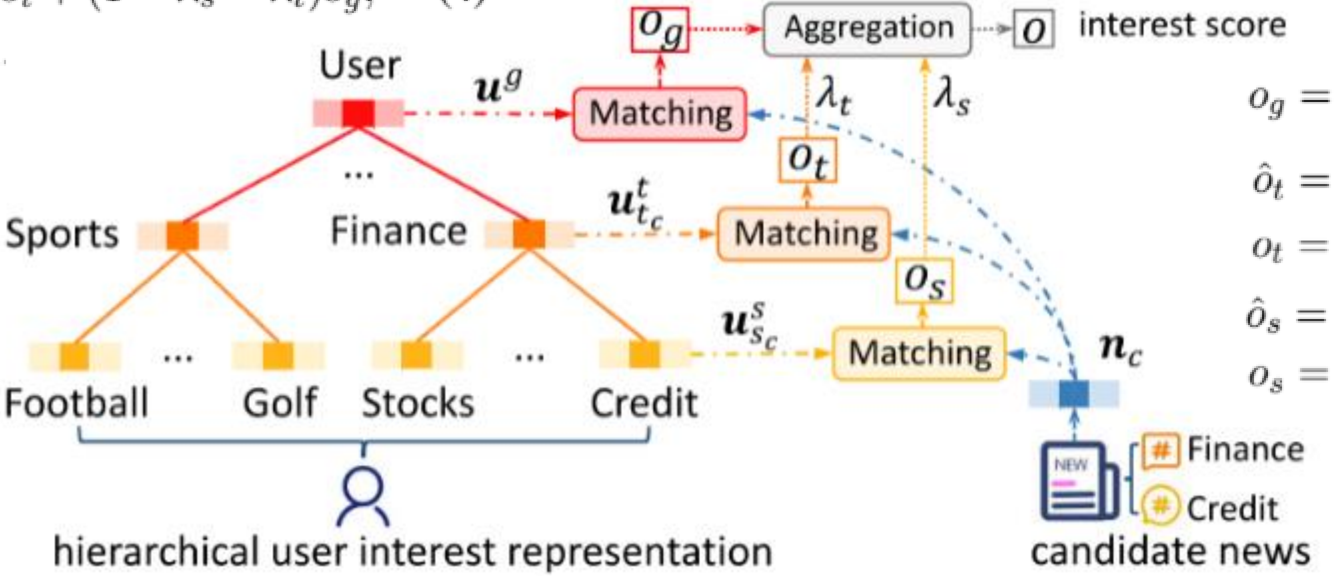
Hierarchical User Interest Matching

$$o = \lambda_s o_s + \lambda_t o_t + (1 - \lambda_s - \lambda_t) o_g, \qquad (4)$$

$$\lambda_t + \lambda_s < 1$$

λ_t = 0.7

λ_s = 0.15



$$o_g = \mathbf{n}_c \cdot \mathbf{u}^g.$$

$$\hat{o}_t = \mathbf{n}_c \cdot \mathbf{u}^t_{t_c}$$

$$o_t = \hat{o}_t * w_{t_c}$$

$$\hat{o}_s = \mathbf{n}_c \cdot \mathbf{u}^s_{s_c}$$

$$o_s = \hat{o}_s * w_{s_c}$$

Loss function

- NCE loss
- Given a positive sample n_i^+ (a clicked news) in the training dataset *O*,
- random select *K* negative samples $[n_i^1, ..., n_i^K]$ (non-clicked news)

$$\mathcal{L} = -\sum_{i=1}^{|\mathcal{O}|} \log \frac{\exp(o_i^+)}{\exp(o_i^+) + \sum_{j=1}^{K} \exp(o_i^j)}. \quad (5)$$

Outline

Datasets

	# News	# Topics	# Subtopics	# Users	# Clicks
MIND	65,238	18	270	94,057	347,727
Feeds	1,126,508	28	-	50,605	473,697

Performance

	MIND				*Feeds*			
	AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
EBNR	61.62±0.15	28.07±0.18	30.55±0.22	37.07±0.21	63.48±0.32	28.01±0.18	32.05±0.23	37.64±0.22
DKN	63.99±0.23	28.95±0.08	31.73±0.14	38.38±0.17	62.94±0.22	28.05±0.26	32.15±0.34	37.68±0.36
DAN	64.68±0.13	29.78±0.13	32.63±0.21	39.27±0.15	62.67±0.49	27.75±0.34	31.74±0.44	37.42±0.43
NAML	64.30±0.30	29.81±0.17	32.64±0.24	39.11±0.20	64.48±0.24	28.99±0.13	33.37±0.16	38.90±0.18
NPA	64.28±0.53	29.64±0.33	32.28±0.37	38.93±0.39	64.02±0.63	28.71±0.39	33.01±0.50	38.55±0.47
LSTUR	65.68±0.35	30.44±0.39	33.49±0.45	39.95±0.39	65.01±0.13	29.28±0.06	33.74±0.09	39.16±0.11
NRMS	65.43±0.15	30.74±0.18	33.13±0.17	39.66±0.15	65.27±0.19	29.40±0.15	33.89±0.16	39.34±0.15
KRED	65.89±0.31	30.80±0.32	33.78±0.27	40.23±0.26	65.51±0.11	29.57±0.06	34.04±0.06	39.60±0.05
GNewsRec	65.91±0.21	30.50±0.21	33.56±0.21	40.13±0.18	65.23±0.16	29.36±0.11	33.87±0.13	39.44±0.12
FIM	64.65±0.14	29.70±0.17	32.51±0.25	39.30±0.16	65.41±0.23	29.57±0.18	34.08±0.25	39.56±0.23
HieRec	**67.95**±0.14	**32.87**±0.08	**36.36**±0.07	**42.53**±0.10	**66.23**±0.10	**29.82**±0.11	**34.42**±0.13	**39.94**±0.13

(4) NAML

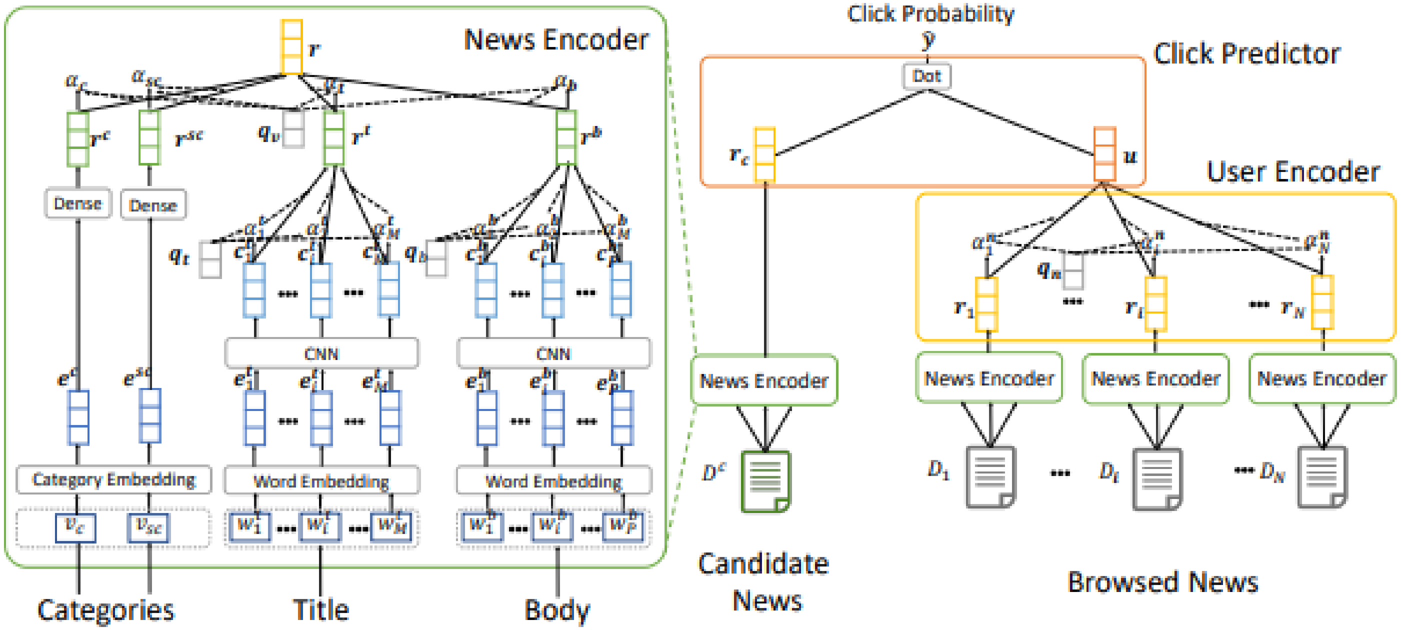


Figure 2: The framework of our *NAML* approach for news recommendation.

(5)NPA

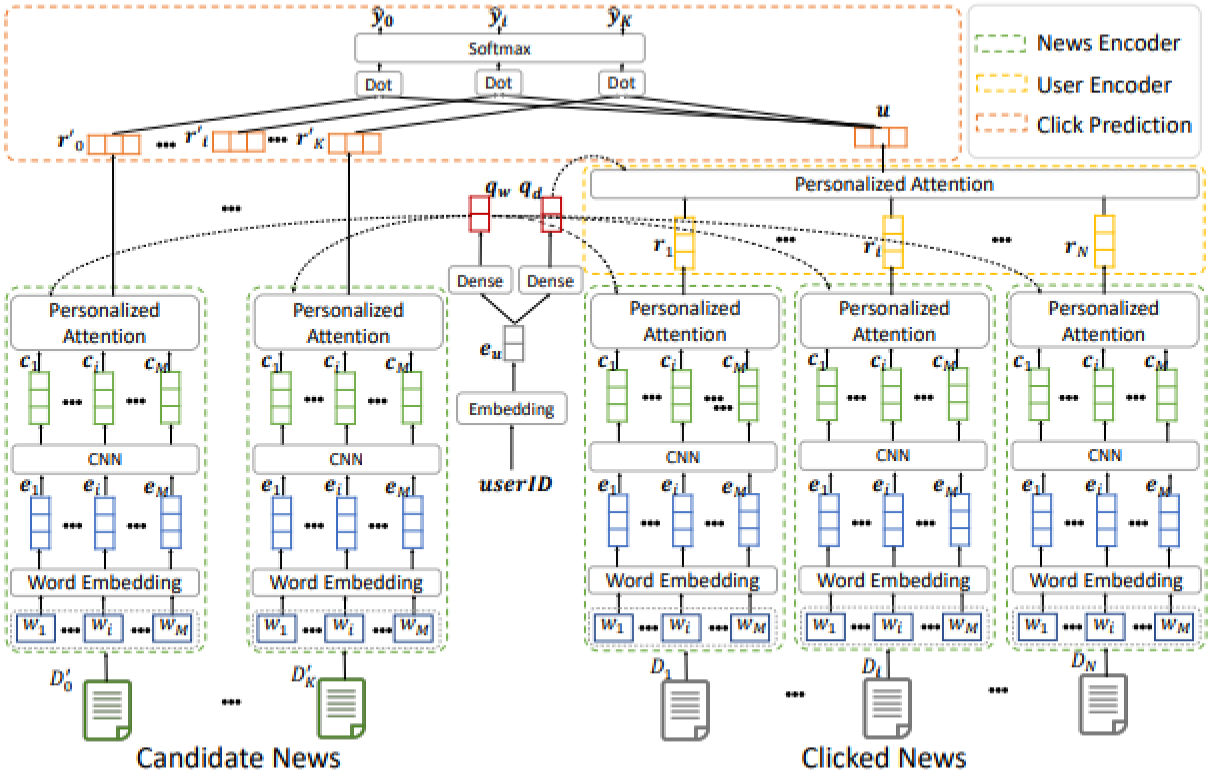


Figure 2: The framework of our NPA approach for news recommendation.

(7) NRMS

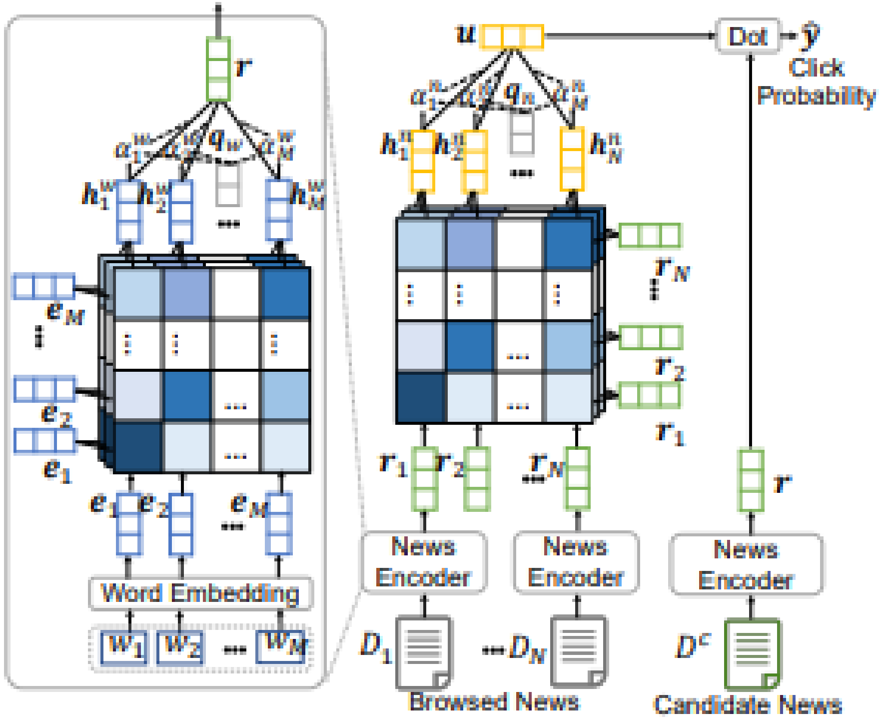


Figure 2: The framework of our *NRMS* approach.

(10) FIM

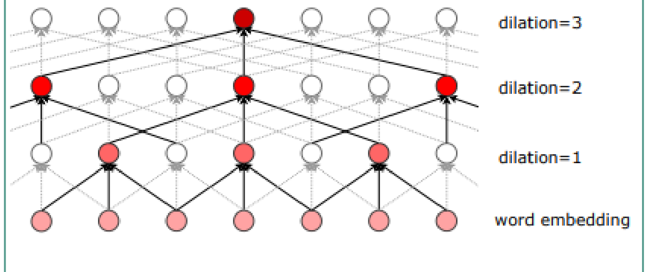


Figure 3: Hierarchical Dilated Convolution (HDC).

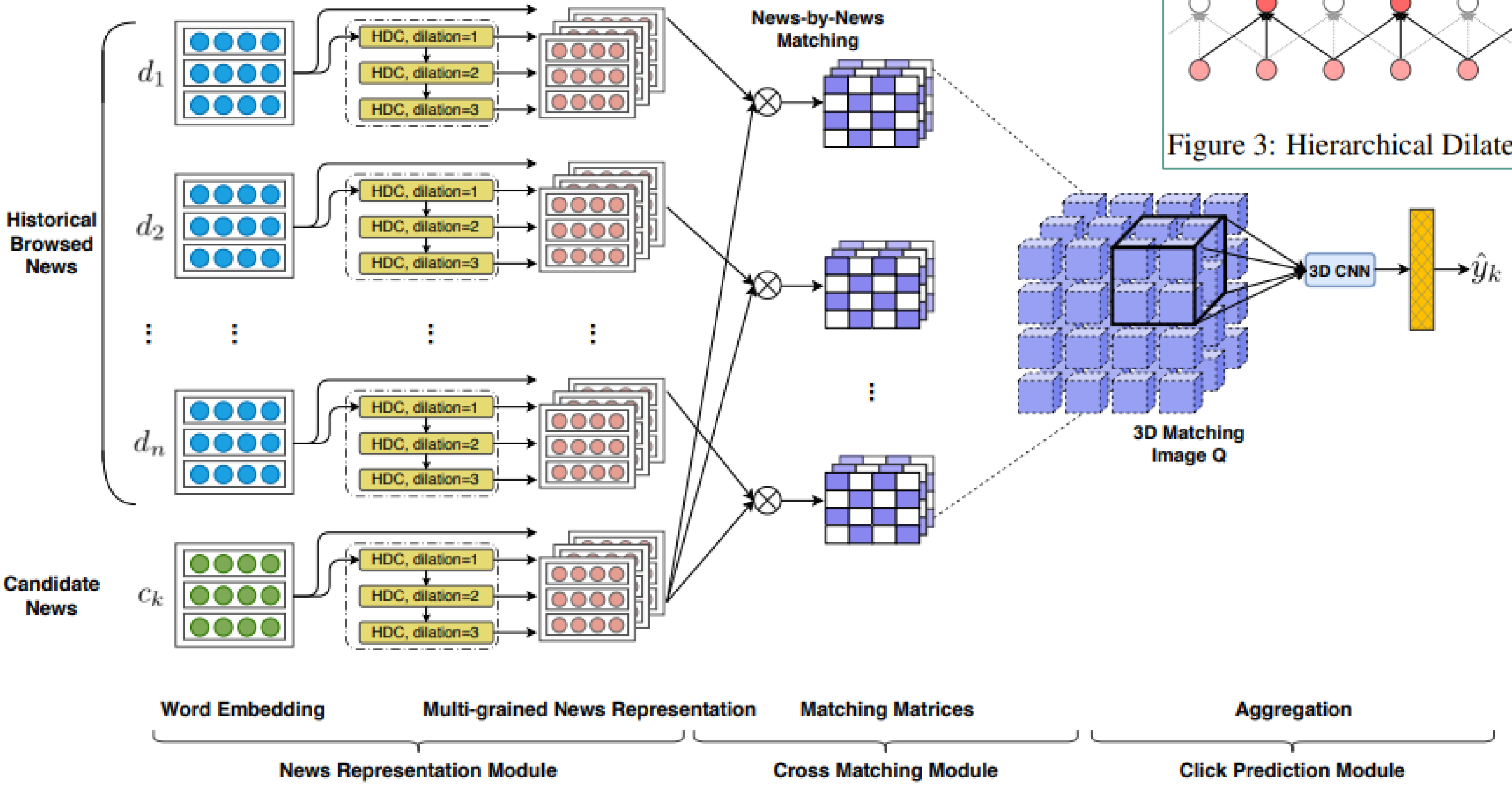
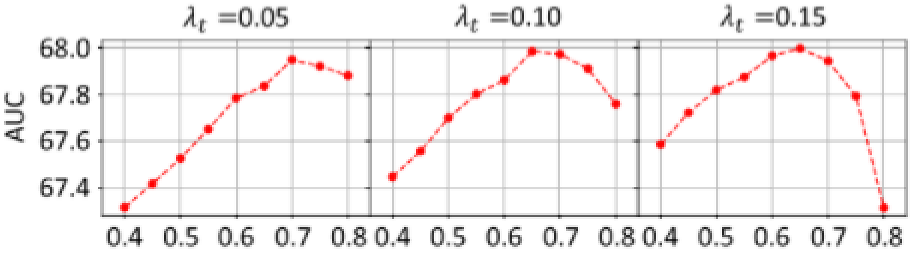


Figure 2: Architecture of our FIM model. *HDC (hierarchical dilated convolution)* is the news encoder.

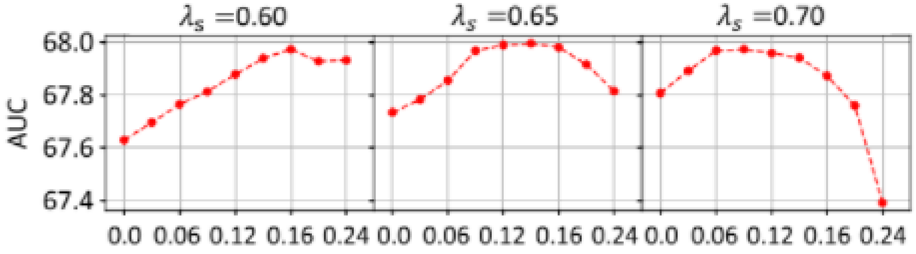
Hyper-parameters

λ_t = 0.7



(a) Performance of *HieRec* with different λ_s

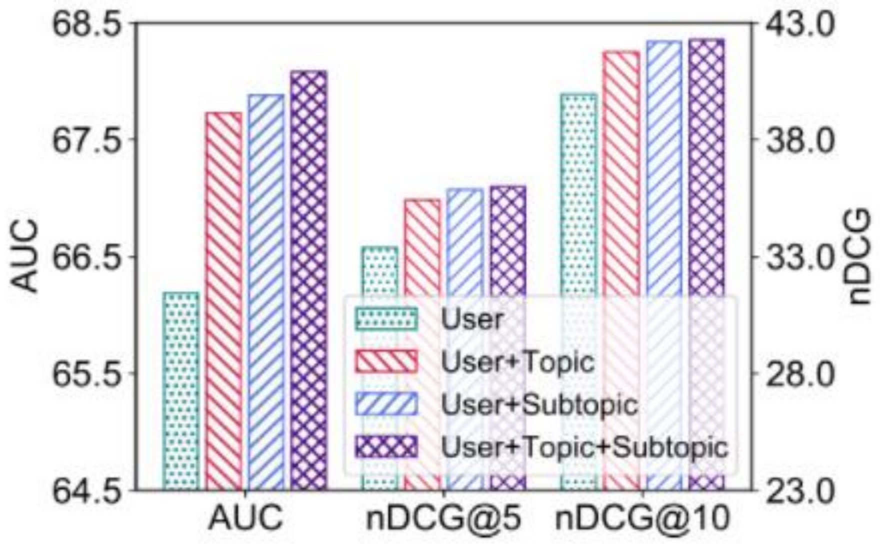
λ_s = 0.15



(b) Performance of *HieRec* with different λ_t

$$o = \lambda_s o_s + \lambda_t o_t + (1 - \lambda_s - \lambda_t) o_g, \qquad (4)$$

Effect of hierarchical user interest modeling



Outline

Conclusion

- Propose a news recommendation method with hierarchical user interest modeling, named HieRec.
- HieRec is a three-level hierarchy to represent user interest for matching different aspects and granularity.