# Accurate News Recommendation Coalescing Personal and Global Temporal Preferences

**Bonhun Koo, Hyunsik Jeon, U Kang**
*Seoul National University*

## PAKDD 2020

# Goal

- To discuss
  - News recommendation on online news service
  - News data patterns
    - *Popularity/Freshness* patterns
  - News Recommendation Coalescing Personal and Global Temporal Preferences (PGT)
    - How well does PGT exploit news data patterns to provide accurate news recommendation?

# Outline

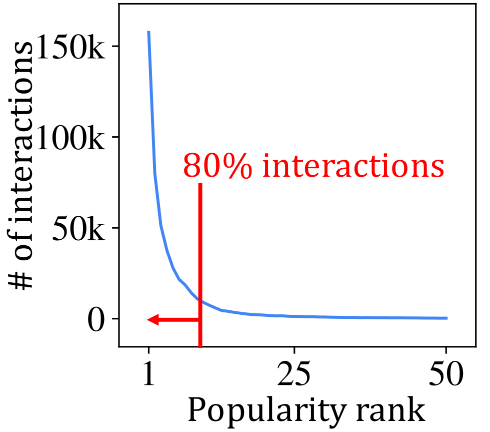
- **Introduction**
- Proposed Method
- Experiments
- Conclusion

# Online News Service

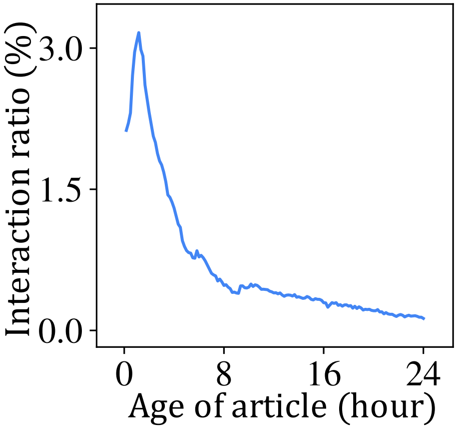
- Online news service
  - Thousands of news everyday
  - Millions of users
- Challenges
  - **Newly published** news articles everyday!
    - The cold-start problem
  - News recommendation considering users' interests
    - Individual personal preference
    - Time-dependent preference

# Popularity/Freshness patterns

- News data patterns
  - Popularity pattern
    - Users mostly prefer popular news
  - Freshness pattern
    - # of interactions of news rapidly decreases over age




(a) Popularity pattern




(b) Freshness pattern

# Problem Definition

- **Input**
  - News watch history of each user $u$
  - Candidate news articles at time $t$
  - Contents of news article
- **Output**
  - Ranks of candidates for each user $u$ at time $t$
- **Such that**
  - The cold-start problem
  - Popularity/Freshness patterns of news

# Outline

- Introduction
- **Proposed Method**
- Experiments
- Conclusion

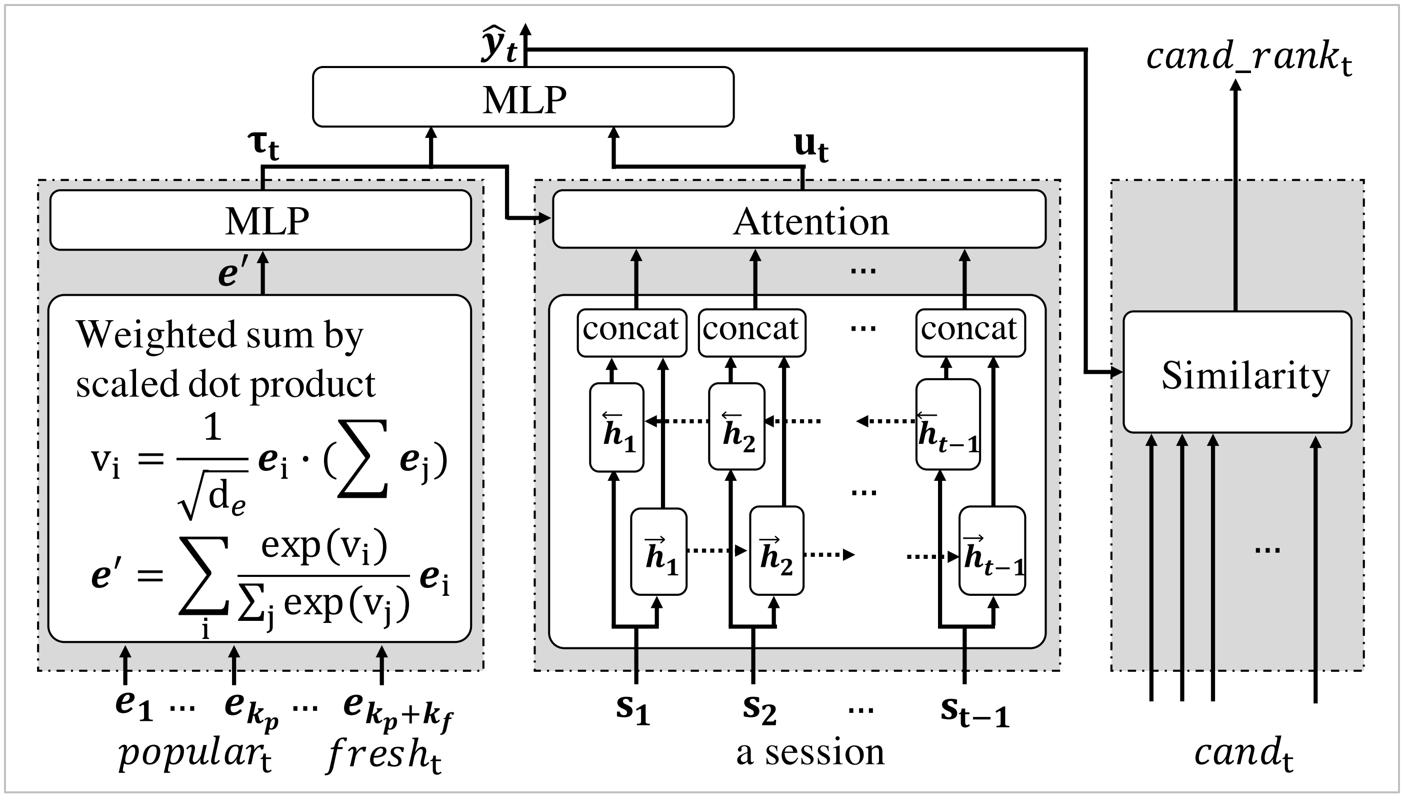
# Proposed Method

- **PGT** (News Recommendation Coalescing **P**ersonal and **G**lobal **T**emporal Preferences)

- **Main intuition**
  - Global temporal preference
    - Comprehensive preference of all users at recommendation time
  - Attention network for the personal preference
    - To deal with a quick change of personal preference
    - The global temporal preference vector is used as context

# Proposed Method

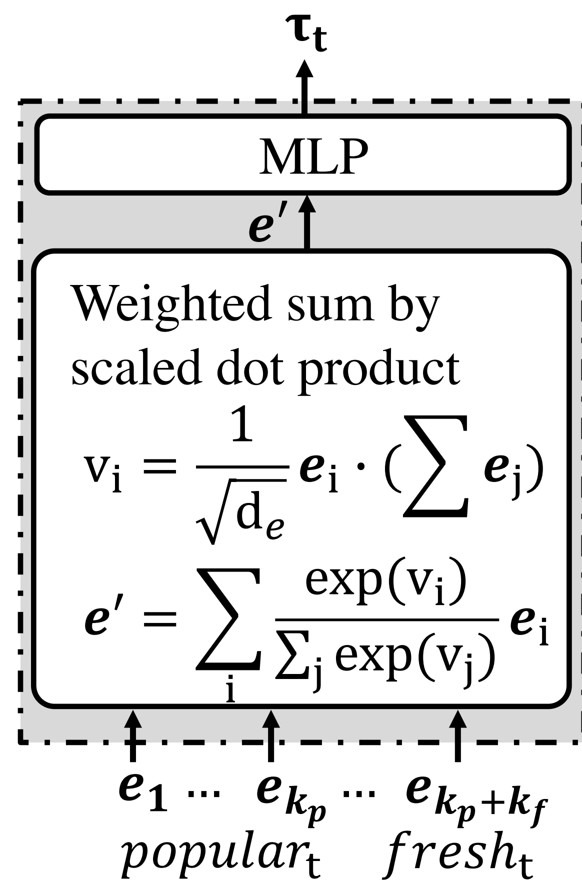
- Overview of PGT




(a) Global temporal preference (b) Personal preference (c) Ranking candidates

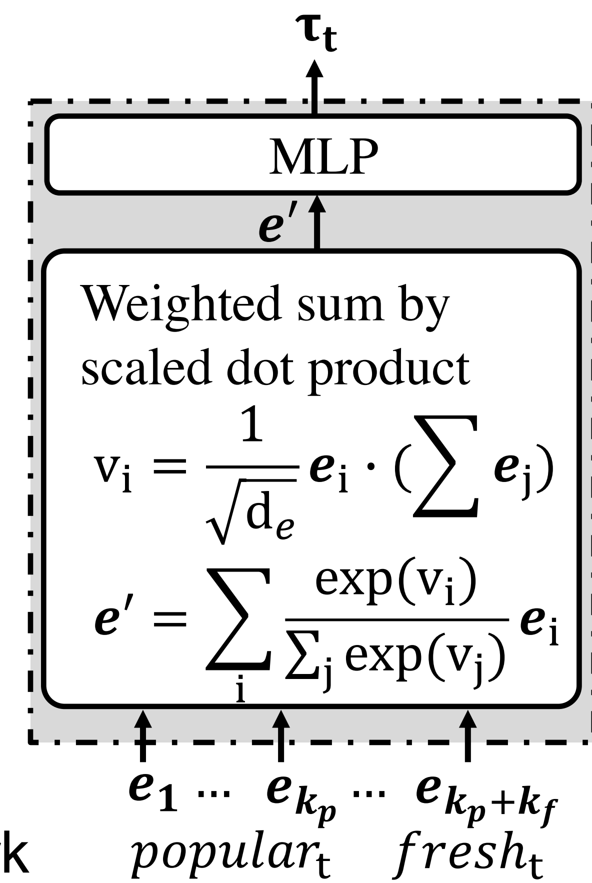
# Proposed Method

- Global temporal preference
  - **Intuition**
    - Comprehensive preference of all users at recommendation time $t$
    - Extract time-dependent features
      - To deal with popularity and freshness patterns
    - To recommend newly published articles well
      - To better handle the cold-start problem



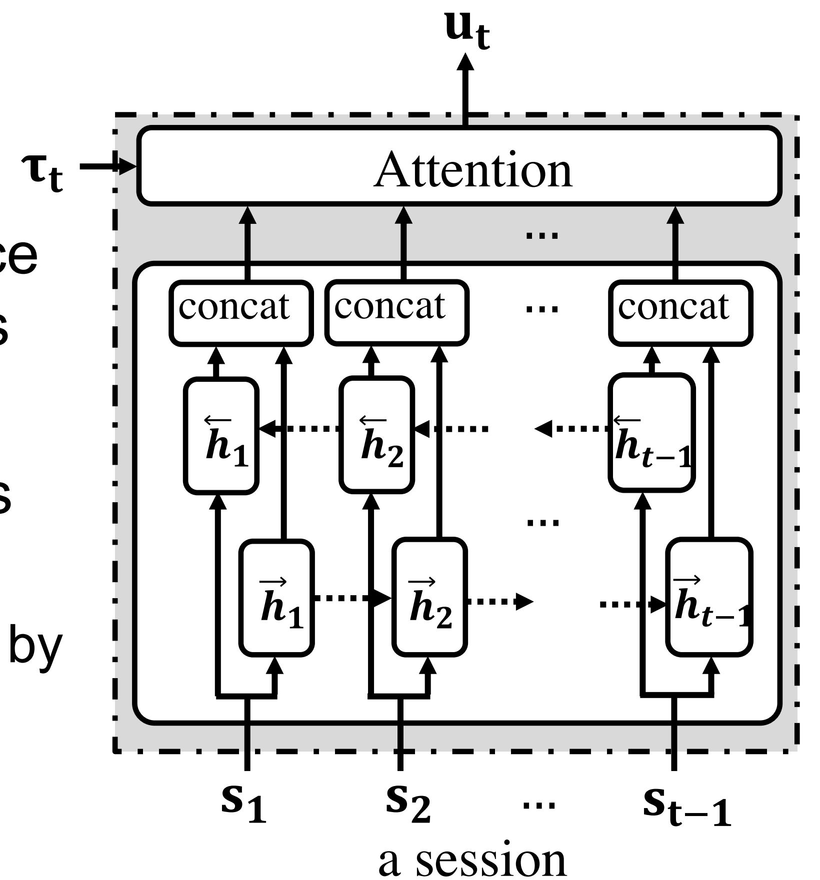
# Proposed Method

- Global temporal preference
  - **Input**
    - $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{k_p}, \dots, \boldsymbol{e}_{k_p+k_f}$
      - Popular/Fresh articles
      - $k_p$: # of popular articles
      - $k_f$: # of fresh articles
  - **Output**
    - $\boldsymbol{\tau}_t$: Global temporal preference
  - **How**
    - Weighted sum by attention network
      - $v_i$: unnormalized attention score



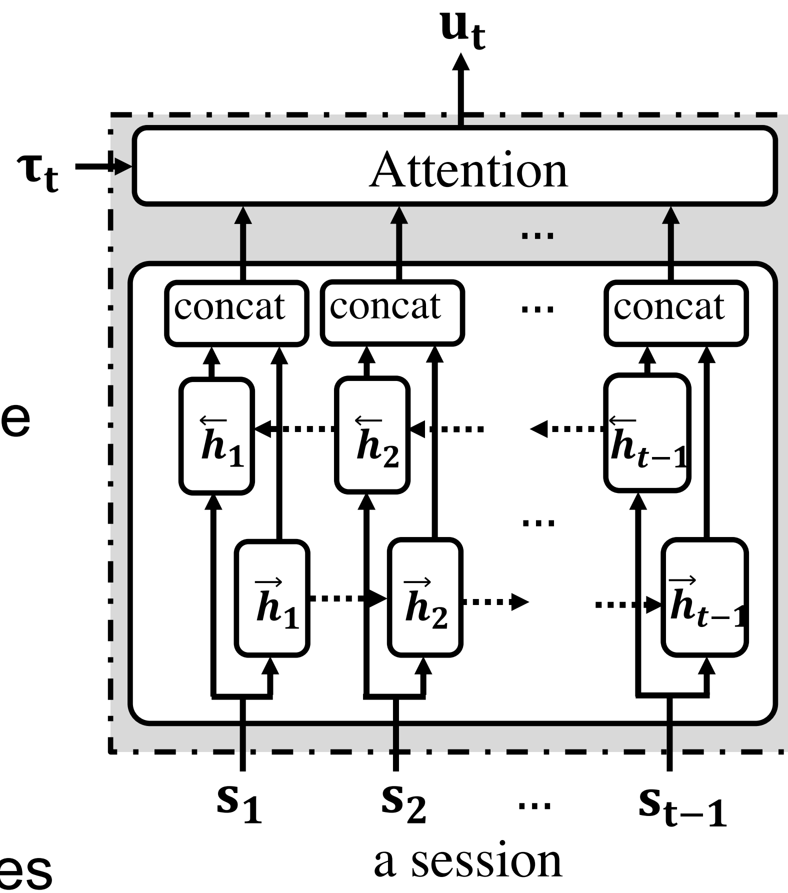
# Proposed Method

- Personal preference
  - **Intuition**
    - Individual personal preference from previous user behaviors
    - Highlight important behaviors using the attention network
      - Time-dependent highlighting by $\boldsymbol{\tau}_t$



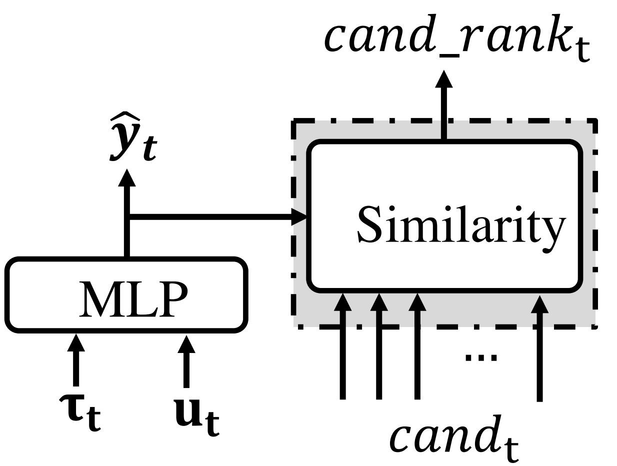
# Proposed Method

- Personal preference
  - **Input**
    - $s_1, s_2, \dots, s_{t-1}$
      - Previous watches of a user
    - $\tau_t$: global temporal preference
  - **Output**
    - $u_t$: personal preference
  - **How**
    - Bidirectional RNN
    - Weighted sum of hidden states
      - By attention network using $\tau_t$ as context



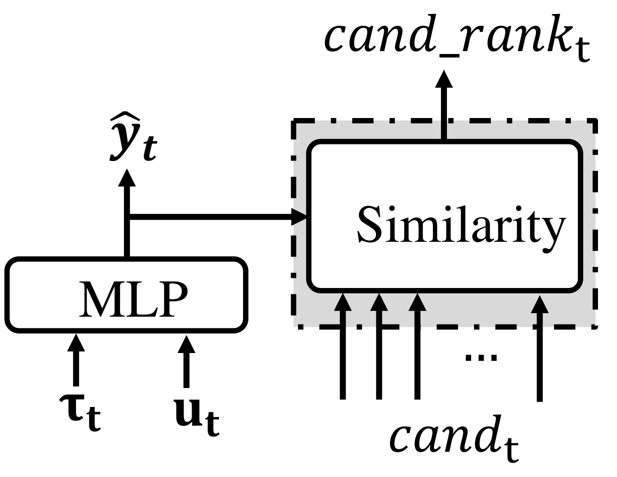
# Proposed Method

- Ranking candidates
  - **Intuition**
    - Generate prediction vector $\hat{\boldsymbol{y}}_t$ from two preferences $\boldsymbol{\tau}_\mathrm{t}$, and $\mathbf{u}_\mathrm{t}$
    - Scores each candidate articles by utilizing $\hat{\boldsymbol{y}}_t$, then ranks candidates
  - **Similarity**
    - Inverse of L2 distance between $\hat{\boldsymbol{y}}_t$ and candidate article vector



# Proposed Method

- Ranking candidates
  - **Input**
    - $\boldsymbol{\tau}_t$: global temporal preference
    - $\boldsymbol{u}_t$: personal preference
    - $cand_t$: candidate articles
  - **Output**
    - $cand\_rank_t$: rank of candidates
  - **How**
    - Measure the similarity between prediction vector $\hat{\boldsymbol{y}}_t$ and candidate article vector



# Outline

- Introduction
- Proposed Method
- **Experiments**
- Conclusion

# Experimental Question

- Q1. **Accuracy** on news recommendation
- Q2. **Effect of m**odeling the **global temporal preference**
- Q3. **Effect of** modeling the **attention network** in modeling personal preference

# Datasets

- Datasets
  - Adressa: user-news interaction of 'Adresseavision' in Norway
  - Globo: user-news interaction of 'G1' in Brazil
- Summary of datasets

| Dataset | # Sessions | # Events | # Articles | Period |
|---|---|---|---|---|
| Adressa 1W[1] | 112,405 | 487,961 | 11,069 | 7 days |
| Adressa 10W[1] | 655,790 | 8,167,390 | 43,460 | 90 days |
| Globo[2] | 296,332 | 2,994,717 | 46,577 | 16 days |

[1]: http://reclab.idi.ntnu.no/dataset
[2]: https://www.kaggle.com/gspmoreira/news-portal-user-interactions-by-globocom

# Competitors

- Competitors
  - Only popularity
    - POP
  - RNN-based method
    - Park et al. [CIKM'17]
    - Okural et al. [SIGKDD'17]
  - 3-D CNN method
    - Weave&Rec [Khattar et al. CIKM'18]
  - Attention-based method
    - HRAM [Khattar et al. CIKM'18]
    - NPA [Wu, C. et al. SIGKDD'19]

# Experimental Setup

- Training method
  - Divide data into training, validation, and test sets with ratio of 8:1:1 based on the interaction time
  - To maximize the similarity between 1) a prediction vector 2) and the corresponding selected article vector
    - PGT
      - Loss function: mean squared error (MSE) of two vector
      - Optimizer: Adam optimizer
    - Competitors: follow their best setting
  - Mini-batched inputs of size 512

# Experimental Setup

- Metric
  - HR@5: Hit Rate
  - MRR@20: Mean Reciprocal Rank

$$HR@5 = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} |\{r_i | r_i <= 5\}|$$

$$MRR@20 = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} c_i, \qquad c_i = \begin{cases} \frac{1}{r_i}, & \text{if } r_i \leq 20 \\ 0, & \text{otherwise} \end{cases}$$

# Q1. Accuracy

- Q1. How well does PGT recommend news articles?
  - PGT shows the best performance for all datasets

| Dataset | Metric | POP | Park et al. [11] | Okura et al. [10] | Weave&Rec [5] | HRAM [4] | NPA [16] | PGT |
|---|---|---|---|---|---|---|---|---|
| Adressa 1W | HR@5 | 0.4988 | 0.4714 | 0.4569 | 0.4377 | 0.5347 | 0.6512 | **0.8668** |
| | MRR@20 | 0.3291 | 0.3361 | 0.3341 | 0.3013 | 0.3452 | 0.4983 | **0.6857** |
| Adressa 10W | HR@5 | 0.5672 | 0.3677 | 0.3477 | 0.3007 | 0.3941 | 0.5819 | **0.7106** |
| | MRR@20 | 0.3735 | 0.2461 | 0.2320 | 0.2101 | 0.2531 | 0.3818 | **0.6197** |
| Globo | HR@5 | 0.2845 | 0.3551 | 0.3537 | - | 0.4474 | - | **0.5663** |
| | MRR@20 | 0.2001 | 0.2483 | 0.2500 | - | 0.3101 | - | **0.5116** |

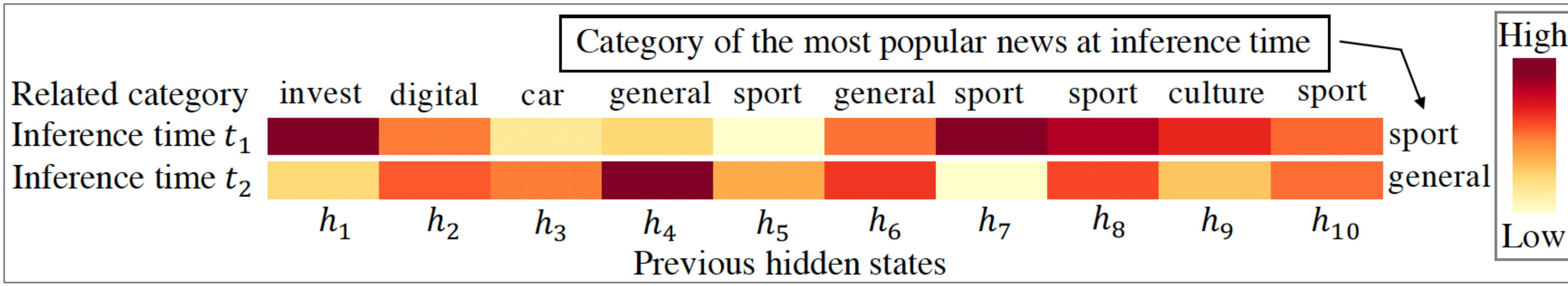
# Q2. Effect of modeling the global temporal preference

- Q2. Does the modeling of **global temporal preference** help improve the accuracy?
  - $PGT_{-T}$: without the global temporal preference
  - $PGT_{-A}$: without the attention network of BiLSTM

| Dataset | Metric | PGT$_{-T}$ | PGT$_{-A}$ | PGT |
|---|---|---|---|---|
| Adressa 1W | HR@5 | 0.6662 | 0.8497 | **0.8668** |
| | MRR@20 | 0.5647 | 0.6756 | **0.6857** |
| Adressa 10W | HR@5 | 0.6360 | 0.6946 | **0.7106** |
| | MRR@20 | 0.5423 | 0.5610 | **0.6197** |
| Globo | HR@5 | 0.5366 | 0.5562 | **0.5663** |
| | MRR@20 | 0.4923 | 0.5035 | **0.5116** |

# Q3. Effect of modeling the attention network in personal preference

- Case study of the attention network




  - **Different attention weights** to the same news watch history when the **inference time is changed**
  - When ‘sport’ or ‘general’ is popular
    - The attention network gives **more weights** to articles in **the same categories**

# Outline

- Introduction
- Proposed Method
- Experiments
- **Conclusion**

# Conclusion

- Proposed **PGT** for recommendation on an online news service
  - To provide accurate recommendation
- **Main idea**: Let's extract time-dependent features by the **global temporal preference**
  - The global temporal preference and attention network in personal preference
    - Better handle the **popularity and freshness patterns** of news
    - Improve the accuracy compared to other competitors

# Thank you !

**Paper Homepage:** https://datalab.snu.ac.kr/pgt/