Embedding-based News Recommendation for Millions of Users

Presenter: Chen Xin

Introduction

Introduction

For general recommendation system:

- ID-based method:
  - Collaborative filtering
  - Low-rank factorization

Introduction

For general recommendation system:

- ID-based method:
  - Collaborative filtering
  - Low-rank factorization

Introduction

- Understanding the content of articles
- Understanding user preferences
- Listing selected articles for individual users based on content and preferences.

Introduction

- Return responses within hundreds of milliseconds

Introduction

- Return responses within hundreds of milliseconds

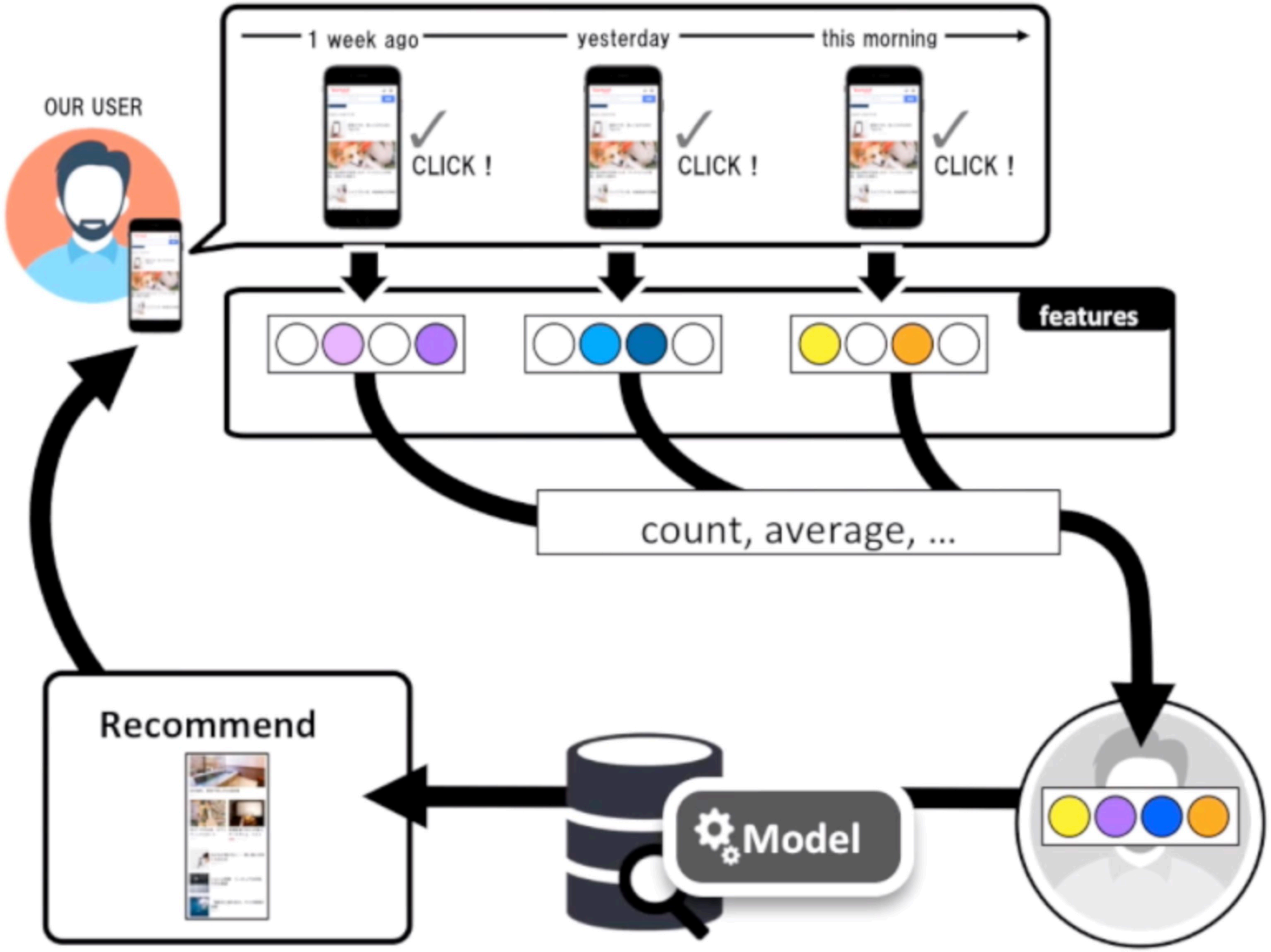
If not

Introduction

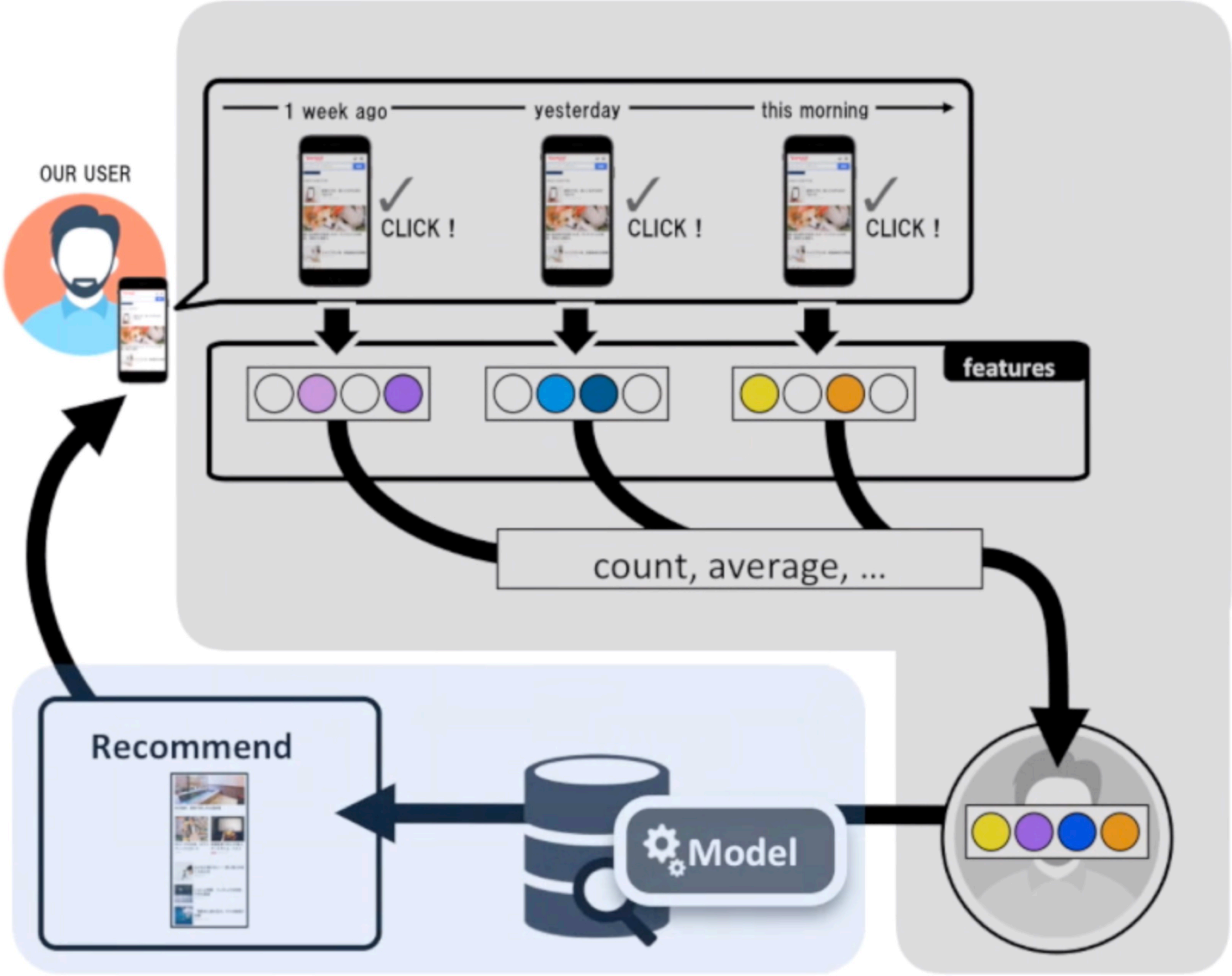
- start with distributed representations of articles based on a variant of a denoising autoencoder.
- generate user representations by using an RNN with browsing histories as input sequences
- Match and list articles for individual users based on inner-product operations by considering system performance.

Example of Yahoo! JAPAN’s homepage on smartphones.

Traditional System



Traditional System

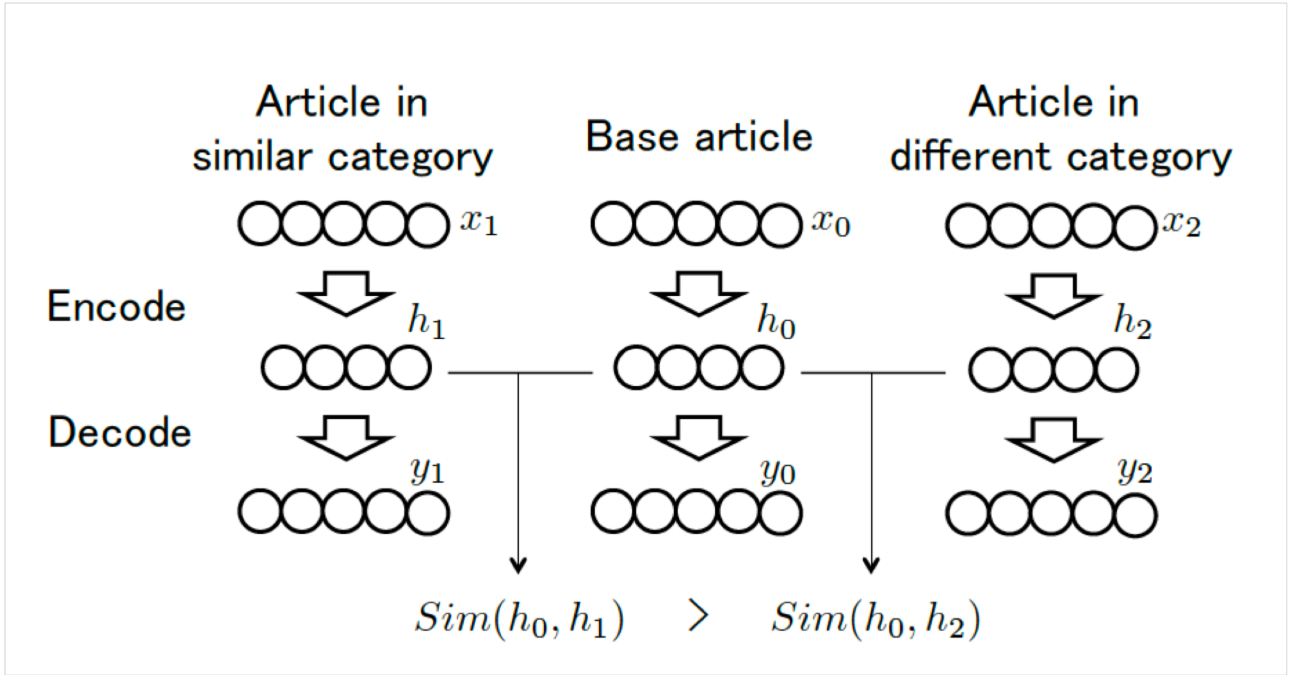


Process Flow

Identify → Matching → Ranking → De-duplication → Advertising

Article Representation

- Generates distributed representation vectors on the basis of a denoising autoencoder (DAE) with weak supervision.



User Representations

Name	Description
BoW	The simplest word-based model
BoW-Ave	Word-based model that uses average function instead of max in Eq.4
BoW-Dec	Word-based model that uses decayed average function with $\beta = 0.9$ similar to that introduced in Section 4.3
Average	Decaying model with $\beta = 1$ (no decaying)
Decay	Decaying model with $\beta = 0.9$
RNN	Recurrent model using simple RNN unit
LSTM	Recurrent model using LSTM-based unit
GRU	Recurrent model using GRU-based unit

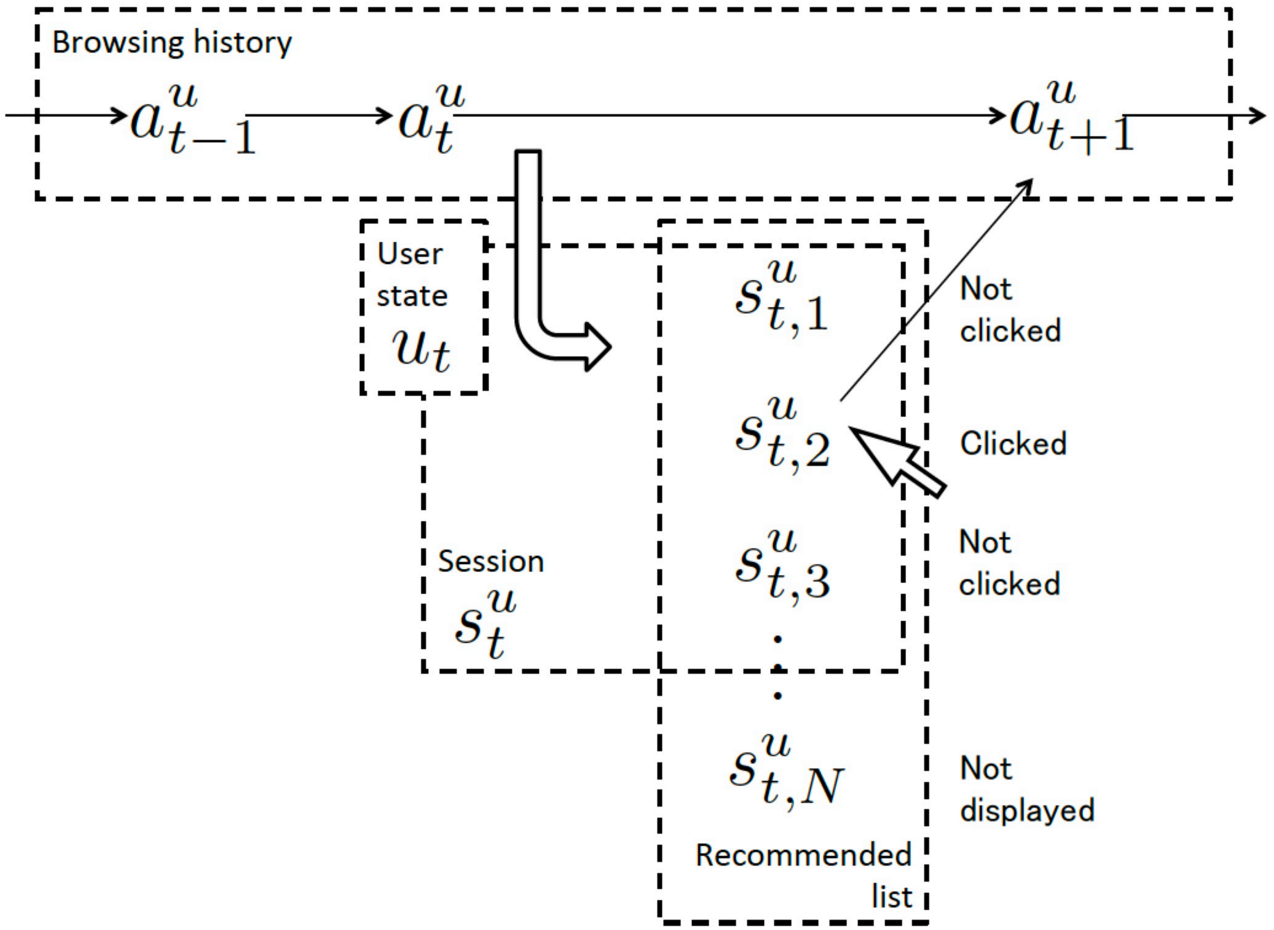
User Representations

Name	Description
BoW	The simplest word-based model
BoW-Ave	Word-based model that uses average function instead of max in Eq.4
BoW-Dec	Word-based model that uses decayed average function with $\beta = 0.9$ similar to that introduced in Section 4.3
Average	Decaying model with $\beta = 1$ (no decaying)
Decay	Decaying model with $\beta = 0.9$
RNN	Recurrent model using simple RNN unit
LSTM	Recurrent model using LSTM-based unit
GRU	Recurrent model using GRU-based unit

User Representations

Name	Description
BoW	The simplest word-based model
BoW-Ave	Word-based model that uses average function instead of max in Eq.4
BoW-Dec	Word-based model that uses decayed average function with $\beta = 0.9$ similar to that introduced in Section 4.3
Average	Decaying model with $\beta = 1$ (no decaying)
Decay	Decaying model with $\beta = 0.9$
RNN	Recurrent model using simple RNN unit
LSTM	Recurrent model using LSTM-based unit
GRU	Recurrent model using GRU-based unit

User Representations

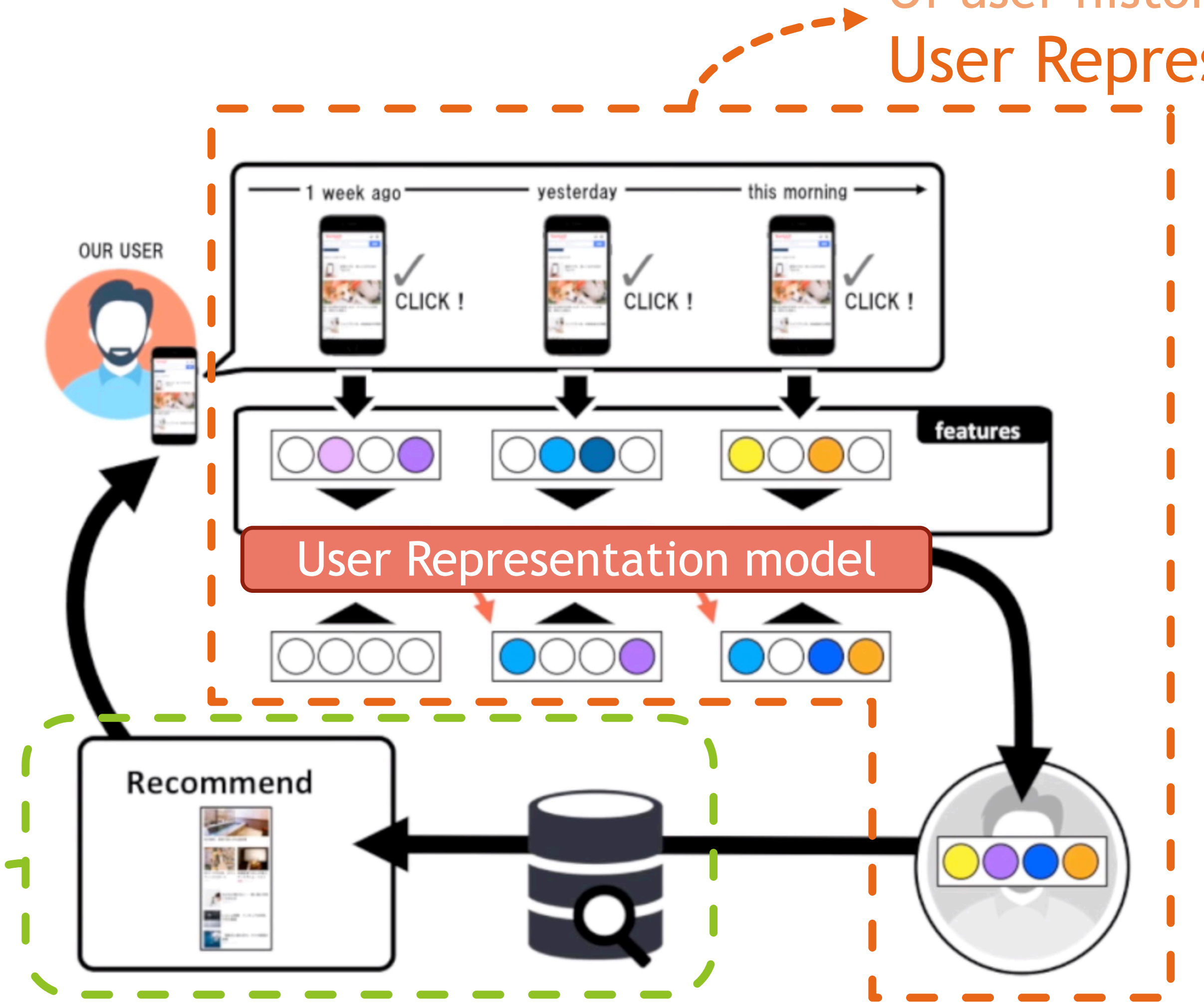


Calculate relevance between user and article

$$\sum_{s_t^u} \sum_{\substack{p_+ \in P_+ \\ p_- \in P_-}} -\frac{\log(\sigma(R(u_t, s_{t,p_+}^u) - R(u_t, s_{t,p_-}^u)))}{|P_+||P_-|}$$

Improvement system

More expressive aggregation
Of user histories using
User Representation model



Simple implementation
Using only the inner products

Dataset

- Training Dataset: 166 million sessions, one billion browses and two million unique articles.
- Test Dataset: 50 million sessions, 20 articles position per person

Offline experimental result

	AUC	MRR	nDCG
BoW	0.582 ± 0.003	0.300 ± 0.003	0.446 ± 0.002
BoW-Ave	0.579 ± 0.004	0.310 ± 0.003	0.452 ± 0.002
BoW-Dec	0.560 ± 0.004	0.297 ± 0.004	0.442 ± 0.003
Average	0.608 ± 0.003	0.313 ± 0.003	0.457 ± 0.002
Decay	0.596 ± 0.003	0.302 ± 0.002	0.449 ± 0.001
RNN	0.612 ± 0.004	0.309 ± 0.004	0.455 ± 0.003
LSTM	0.648 ± 0.004	0.344 ± 0.004	0.481 ± 0.003
GRU	**0.652 ± 0.003**	**0.347 ± 0.004**	**0.484 ± 0.003**

Online experimental result

Metric	ALL	Heavy	Medium	Light
Sessions	+2.3%	+1.1%	+1.0%	+1.8%
Duration	+7.8%	+4.9%	+13.3%	+17.4%
Clicks	+19.1%	+14.3%	+26.3%	+42.3%
CTR	+23.0%	+18.7%	+29.8%	+45.1%

- Heavy: User visited > 5 days during previous week.
- Medium: User visited 2~5 days during previous week.
- Light: User visited < 2 days during previous week.

Heavy : Medium : Light= 3 : 2 : 1

Conclusion

- This paper introduces the news recommendation used by Yahoo! Japan for them news mobile application
- They respectively model news articles and users representation and calculate user-article inner-products to sorting the recommended list
- To generating user representation they total used 8 model, through offline experimental, GRU has the best performance in these model.
- Through online experimental, improved system performance is increased by 23% than before

Q&A