Distributed Representations of Web Browsing Sequences for Ad Targeting

Yukihiro Tagami, Hayato Kobayashi, Shingo Ono, Akira Tajima
Yahoo Japan Corporation

Summary of this study

- Apply an NLP approach to obtain user representations
  - Words -> URLs
  - Paragraphs -> Web browsing sequences (as user interests)
- Compare our Web page visits data with Wikipedia data
  - Frequencies of relative position in sequences are significantly different
- On the basis of the analysis, we propose Backward PV-DM
  - Achieved better results on two ad-related data sets

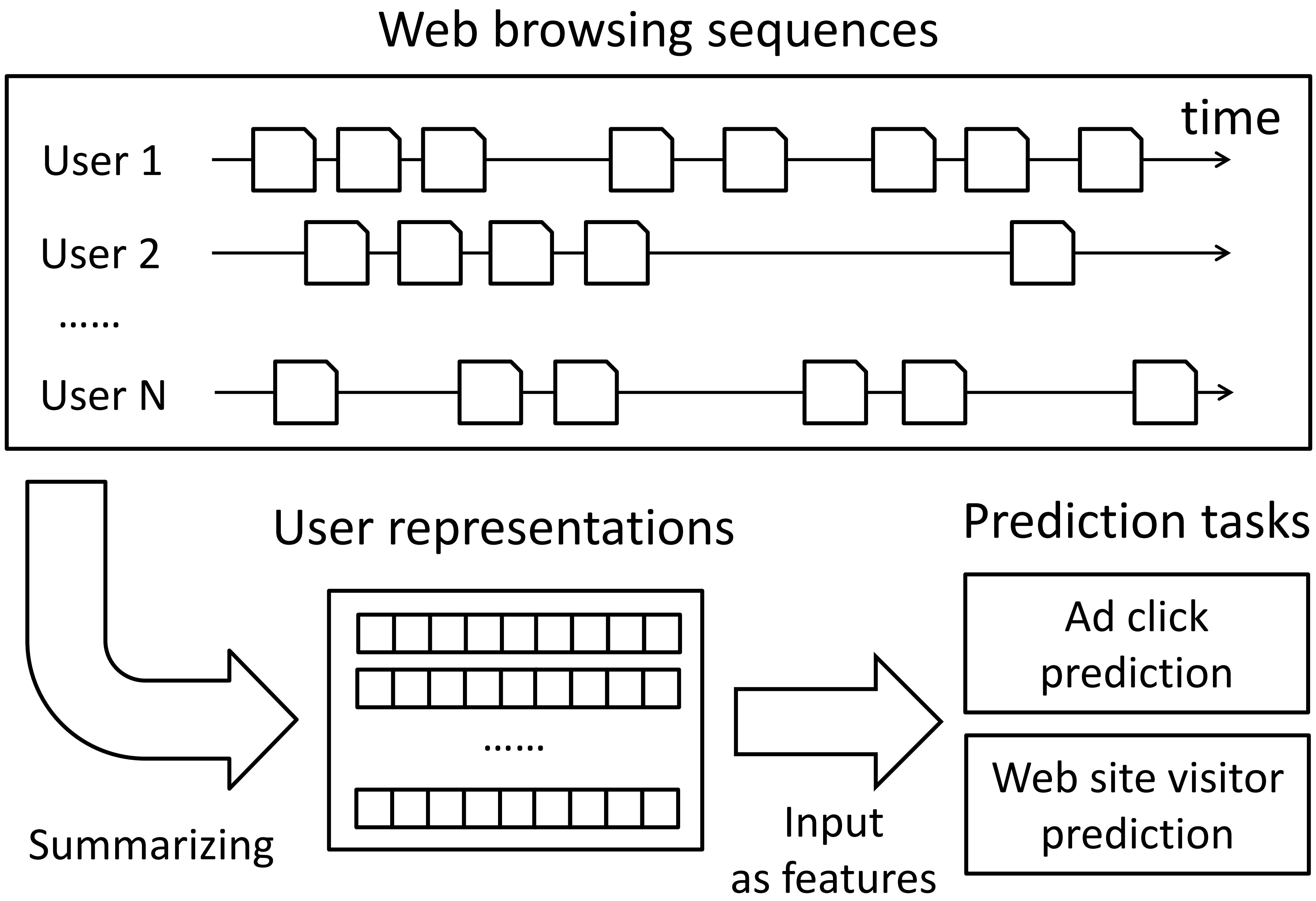
Distributed representations of users from Web page visits

- In our work-in-progress paper, we proposed an approach:
  - To obtain distributed representations of users
  - From Web browsing sequences
  - Using Paragraph Vector

- PV learns distributed representations from pieces of text
  - Words -> URLs
  - Paragraphs -> Web browsing sequences (as user interests)

- Y. Tagami, H. Kobayashi, S. Ono, and A. Tajima. Modeling User Activities on the Web using Paragraph Vector. In *WWW Companion*, 2015.

User representations as features of prediction tasks



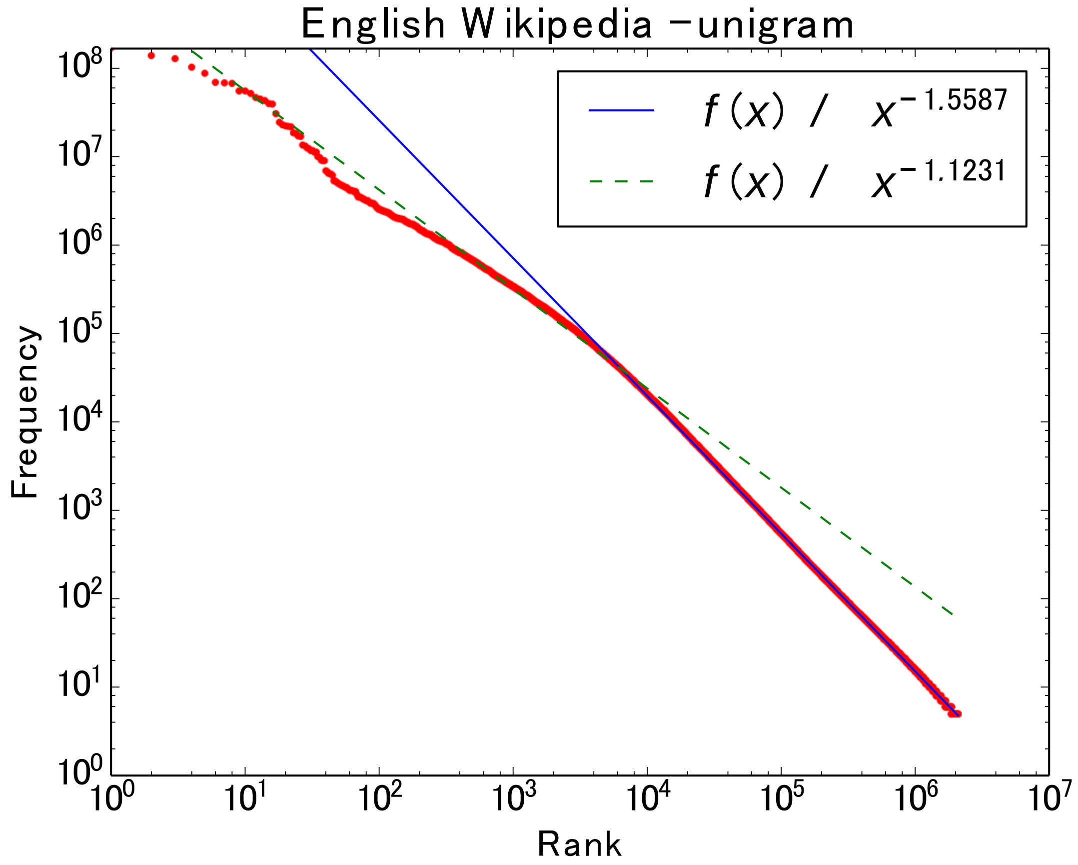
Focusing on the differences of two types of data

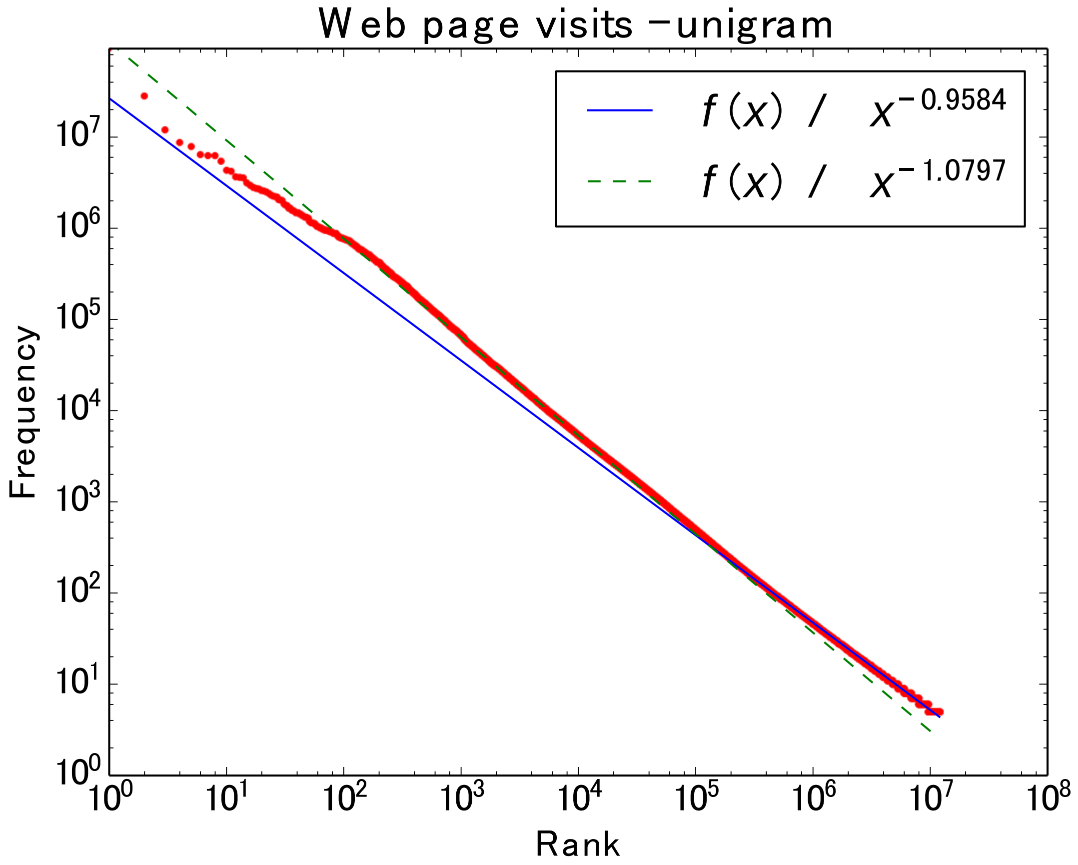
- Two data are probably generated from different distributions
  - Natural language data / Web page visits data

- In this study,
  - We investigate the difference between these distributions
  - On the basis of the difference, we propose Backward PV-DM
  - Evaluate this method on two ad-related prediction tasks

Similarity between two types of data

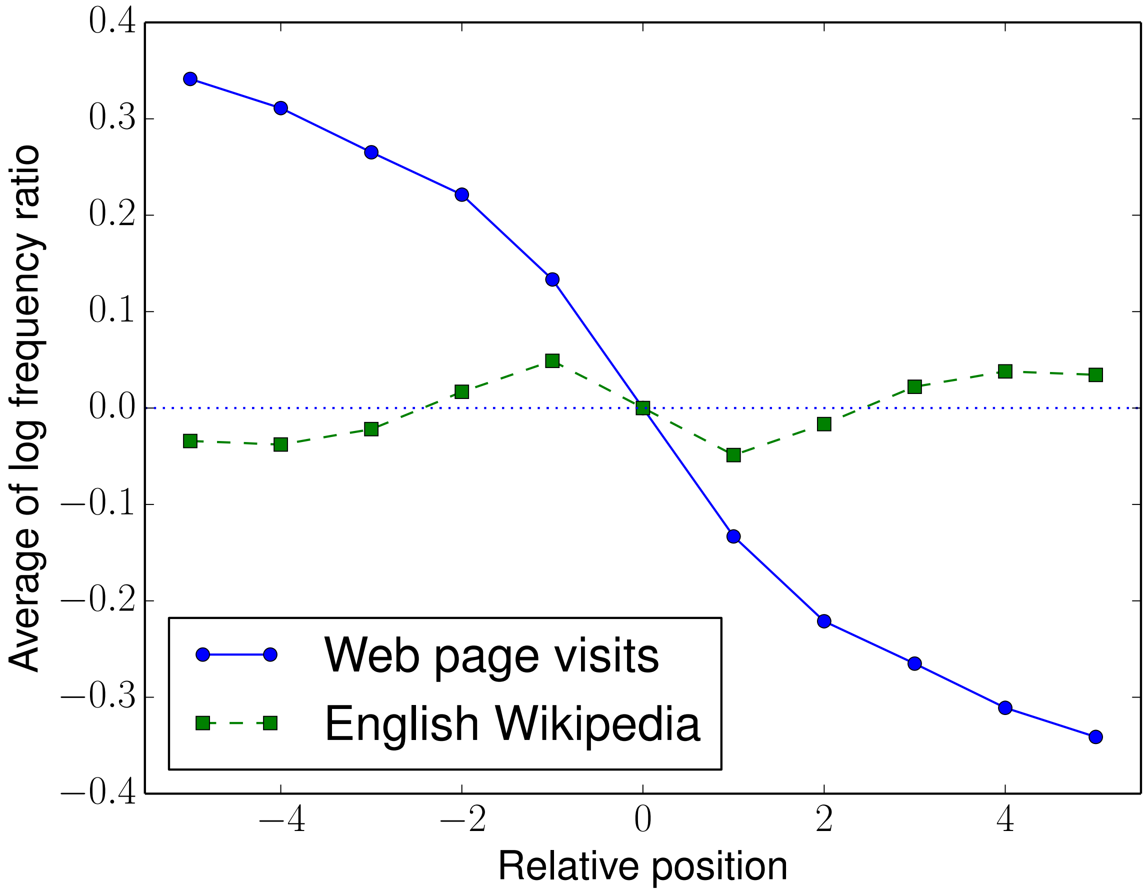
- Both distributions look like roughly straight lines
  - Power-law distribution





Difference between two types of data

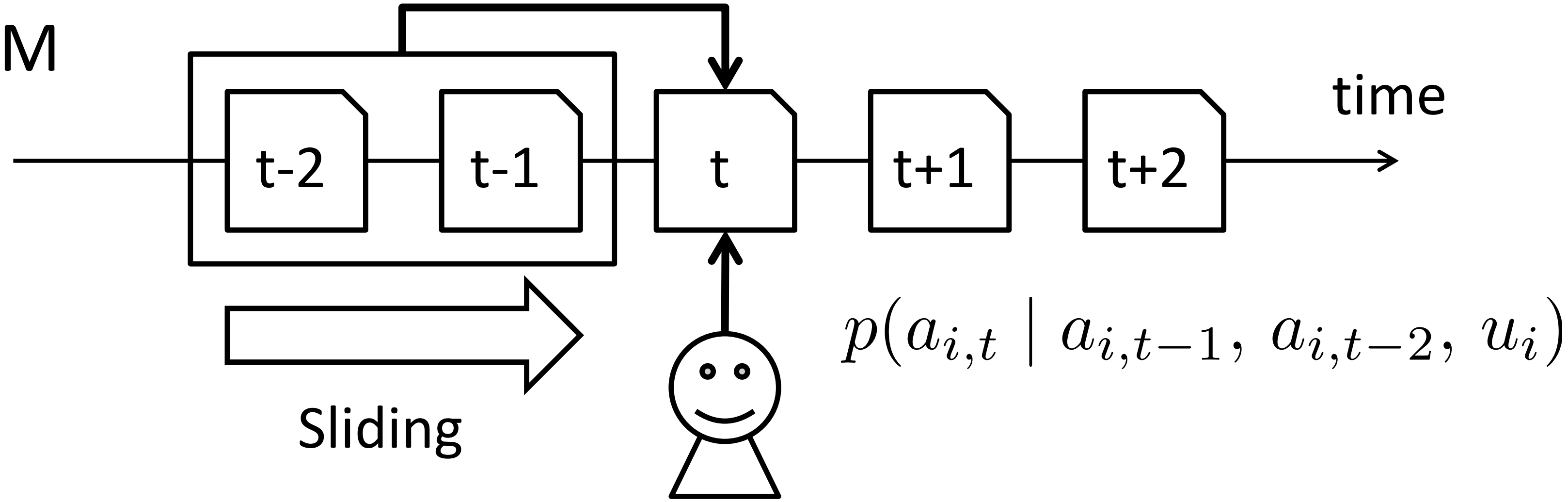
- The "tail" URLs appear in the latter part of a session
- These URLs are important for user modeling



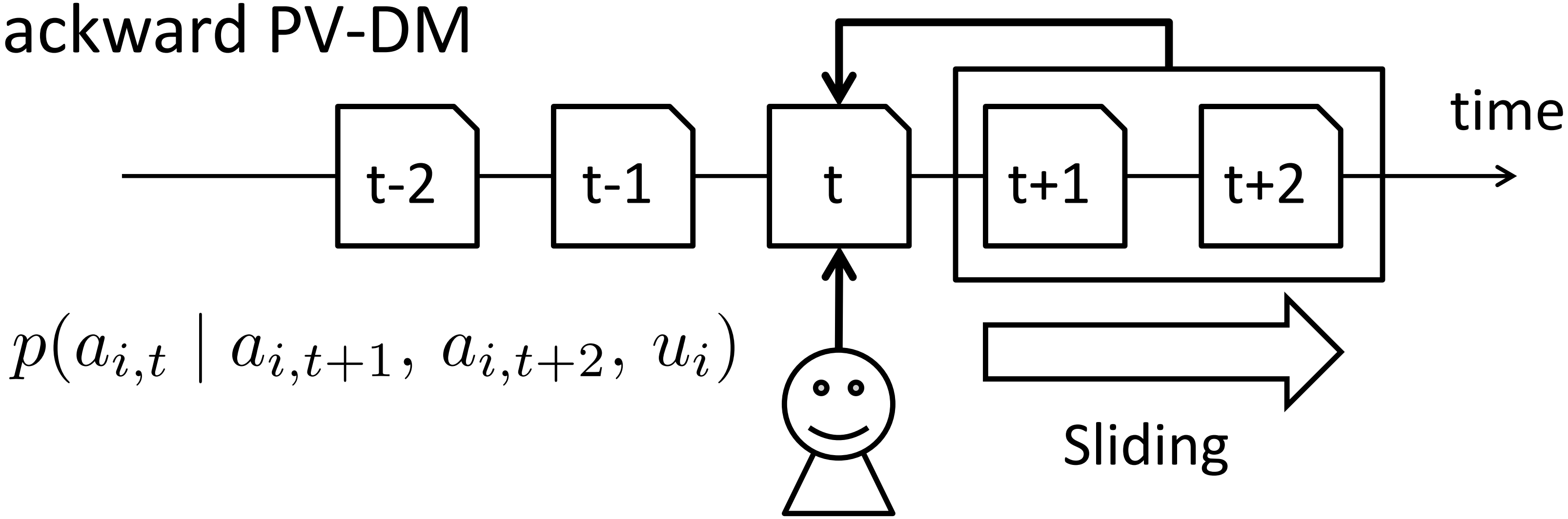
$$\log\left(\frac{\mathrm{freq}(w_{i,t+k})}{\mathrm{freq}(w_{i,t})}\right)$$

The context window is different from the PV-DM

PV-DM



Backward PV-DM



Evaluation settings

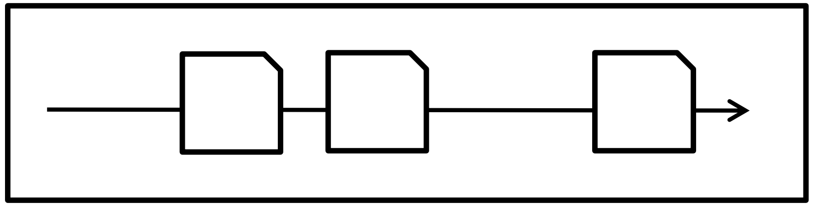
- Two types of ad-related prediction tasks
  - AdClicker
    - Predict clicked contextual ads by each user among five ads
  - SiteVisitor
    - Predict visited advertisers' sites by each user among five sites

- Obtained users' representations using each vector model
  - One task-independent representation for each user
  - One logistic regression classifier for each prediction task

Predicting user's actions from Web browsing history

July 23, 2014

Web browsing sequence of each user

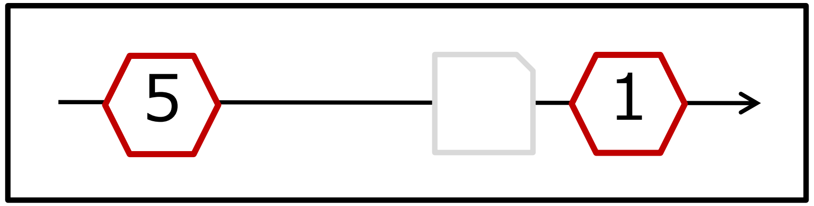


A set of users which selected at least one among five candidates

Predict

July 24, 2014

Labels corresponding to five candidates



Multi-label classification is converted into five binary classification problem

Experimental results

YAHOO! JAPAN

- Using Skip-gram, a user is represented as the simple averaging of vectors of URLs in the sequence
- Backward PV-DM achieved better results than PV-DM

	AdClicker					SiteVisitor				
	Ac1	Ac2	Ac3	Ac4	Ac5	Sv1	Sv2	Sv3	Sv4	Sv5
Skip-gram	**0.9906**	0.8354	**0.6562**	0.7163	**0.7725**	0.8017	0.8328	0.7135	0.7931	0.7417
Directed Skip-gram	0.9904	**0.8374**	0.6533	0.7159	0.7706	0.8019	0.8308	0.7120	0.7914	0.7394
PV-DM	0.9899	0.8151	0.6483	0.7242	0.7633	0.8051	0.8343	0.7180	0.7964	0.7479
Backward PV-DM	0.9902	0.8247	0.6537	**0.7345**	0.7661	**0.8092**	**0.8366**	**0.7222**	**0.8028**	**0.7491**

Values are AUC (Area Under ROC Curve). Larger is better.

Experimental results

- Contextual ads in AdClicker are determined to be displayed by the Web page content as well as user information
- SiteVisitor is the data set based on more complicated user interests

	AdClicker					SiteVisitor				
	Ac1	Ac2	Ac3	Ac4	Ac5	Sv1	Sv2	Sv3	Sv4	Sv5
Skip-gram	**0.9906**	0.8354	**0.6562**	0.7163	**0.7725**	0.8017	0.8328	0.7135	0.7931	0.7417
Directed Skip-gram	0.9904	**0.8374**	0.6533	0.7159	0.7706	0.8019	0.8308	0.7120	0.7914	0.7394
PV-DM	0.9899	0.8151	0.6483	0.7242	0.7633	0.8051	0.8343	0.7180	0.7964	0.7479
Backward PV-DM	0.9902	0.8247	0.6537	**0.7345**	0.7661	**0.8092**	**0.8366**	**0.7222**	**0.8028**	**0.7491**

Values are AUC (Area Under ROC Curve). Larger is better.

Future work

- Other types of features
  - Search queries and Web page contents
- Other than unsupervised learning
  - Semi-supervised, multi-label or multi-task learning
- Sequence modeling with RNNs (Recurrent Neural Networks)
  - Scalable learning methods for Web scale user data

- Now, we apply LSTM-RNN to user browsing sequences
  - For news article recommendation on smartphones

Thank you!
Questions?

Please speak clearly and slowly

Yukihiro Tagami
yutagami@yahoo-corp.jp