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)

User representations as features of prediction tasks

Web browsing sequences

User 1

User 2

......

User N

time

User representations

Prediction tasks

Ad click prediction

Web site visitor prediction

Summarizing

Input as features
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.
The context window is different from the PV-DM

PV-DM

\[ p(a_{i,t} \mid a_{i,t-1}, a_{i,t-2}, u_i) \]

Backward PV-DM

\[ p(a_{i,t} \mid a_{i,t+1}, a_{i,t+2}, u_i) \]
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

July 24, 2014

Labels corresponding to five candidates

Multi-label classification is converted into five binary classification problem
Experimental results

- 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

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

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.

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

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