Background

- Large-scale Web sites that provide various Web services deal with a lot of user-related prediction tasks. However,
  - Some of these tasks have small-scale training data
  - User activities on the Web site are ID-based (such as URLs), and too sparse to be used as features for such cases

- We propose an approach to obtain low-dimensional user vectors from sequences of user activities using recent representation learning approaches in NLP field
  - Considering users and activities as paragraphs and words
  - In this poster, we focus on Web page visits as user activities and represent it as the URL of the Web page

Paragraph Vector

- **PV-DM** (Distributed Memory model of Paragraph Vector)
  - The objective of the vector models for a sequence of i-th user activities is to maximize the sum of log probabilities
    \[
    \sum_{i=1}^{T_i} \log p(a_{i,t} | a_{i,t-1}, \ldots, a_{i,1}, u_i)
    \]
  - The PV-DM defines the probability using the softmax function
    \[
    p(a_{i,t} | a_{i,t-1}, \ldots, a_{i,1}, u_i) = \frac{\exp(w^a_{a_{i,t}}v_{i,t})}{\sum_{a \in A} \exp(w^a_{a_{i,t}}v_{i,t})}
    \]
  - \(w^a_{a_{i,t}}\): “output” vector corresponding to activity \(a_{i,t}\)
  - \(v_{a_{i,t}}\): “input” vector corresponding to activity \(a_{i,t}\)
  - \(v_{u_i}\): “input” vector corresponding to user \(u_i\)
  - \(v_t = [v_{a_{i,t-1}}^{T}, \ldots, v_{a_{i,1}}^{T}, v_{u_i}^{T}]^{T}\): concatenated input vector

Experiments

- **Data sets (prediction tasks)**
  - **AdClicker**
    - Consisting of the users who clicked contextual ads that are included in the five selected ad campaigns (Ac1–Ac5)
  - **SiteVisitor**
    - Consisting of the users who visited Web sites of five selected advertisers (Sv1–Sv5)

<table>
<thead>
<tr>
<th>Training</th>
<th>Validation</th>
<th>Testing</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdClicker</td>
<td>51,576</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>SiteVisitor</td>
<td>1,862,693</td>
<td>20,000</td>
<td>20,000</td>
</tr>
</tbody>
</table>

- We transformed the multi-label problems into a set of binary classification problems and trained logistic regression classifiers
- The evaluation measure is Area Under ROC curve (AUC)

- **Overview of our approach**

  **Sequences of user activities on the Web**

  **User representations**

  **Prediction tasks**

  **Summarizing**

  **Input as features**

  **Learn the vectors to predict the next activity correctly**

- **Methods to extract user representations**
  - Bag of URLs (high-dimensional vectors)
    - Bin: whether the user visited each Web page or not (1/0)
    - Freq: frequencies of the user’s each Web page visits
  - Log-bilinear models (low-dimensional vectors)
    - Skip-gram: simple averaging of activity vectors
    - PV-DM: proposed method using \(v_{a_{i,t}}\)
    - PV-DM+Skip-gram: concatenated vectors of above two

- **Results**
  - PV-DM achieved better results than Skip-gram in SiteVisitor whereas the opposite trend is shown in AdClicker
  - The combination method PV-DM+Skip-gram performed better than individual methods