Using Recurrent Neural Networks to Predict Future Events in a Case with Application to Cyber Security

Article Authors: Stephen Jacob, Dr. Yuansong Qiao, Dr. Paul Jacob, Dr. Brian Lee

Presenter Name: Stephen Jacob

Article ID: 98002

Affiliation: Athlone Institute of Technology (AIT), Software Research Institute (SRI)

Email: s.jacob@research.ait.ie
Stephen Jacob

- B.Sc in Computer Systems at University of Limerick
- M.Sc in Computer Engineering at Athlone Institute of Technology (AIT)
- PhD Student in the Software Research Institute (SRI) at Athlone Institute of Technology (AIT)
Overview

- Cyber attacks target business organizations every day
- Large numbers of security alerts detected
- Cyber security personnel require software support for prioritizing these alerts
  - Predict future events in currently executing business processes
  - Identify critical software services and infrastructure required to enable these events
  - Prioritize security alerts that target these software services
- Primary question: “Can the application of deep learning to process mining be used to predict future events in an ongoing process?”
Paper Overview

- State-of-the-art descriptions for:
  - Process Mining
  - Deep Learning
- Train deep learning networks, specifically LSTM models as an alternative to conventional process mining
- Outline two methods to train LSTMs to predict future events
- Evaluate both methods in terms of accuracy
Process Mining

- The discovery, analysis and modelling of information from logged data.
- Data is comprised of events
- Process model can be extracted from data
- Cases are sequences of events found in model
- ProM is an open-source process mining tool used to generate process models
Deep Learning/Neural networks

- An alternate form of conventional process mining is to
  - train a deep learning neural network model to learn the typical behaviour of cases
  - use model to make predictions about future events

- Neural networks are software systems comprised of a number of layers which output data in response to input received at an input layer

- The neural network used is a **Recurrent Neural Network** (RNN).
Recurrent Neural Networks (RNN)

- Neural network is required to process sequential data
- Events at a single time step of a sequence can affect subsequent events
- RNN model is a neural network where the output is fed back into the model over a number of time steps
- RNNs can be trained on cases
LSTM Model

- Long Short Term Memory (LSTM) is a type of RNN
- Advantages
  - Models noisy, sequential data
  - Processes data with time based fields
  - Detects long-term dependencies
- LSTM network is the hidden layer(s) of our model
- The output layer is a Dense layer where each node outputs a numeric value.
- The actual output is a probability distribution over a number of different event types
- The highest probability value and its corresponding event type is the predicted future event.
## LSTM Model Training Methods

### Teacher Forcing Method

<table>
<thead>
<tr>
<th>X (input)</th>
<th>Y (output)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1, 2, 3, 4, 5]</td>
<td>[2, 3, 4, 5, !]</td>
</tr>
</tbody>
</table>

- Uses the output from the previous time step as input
- For example, along the sequence when predicting the value 2, the input is 1 and so on.
- This method uses a Keras TimeDistributed layer to apply a Dense layer at every time step and output values at every time step.

### Prefix Method

<table>
<thead>
<tr>
<th>X (Input)</th>
<th>Y(Output)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1, 2]</td>
<td>[3]</td>
</tr>
<tr>
<td>[1, 2, 3]</td>
<td>[4]</td>
</tr>
<tr>
<td>[1, 2, 3, 4]</td>
<td>[5]</td>
</tr>
<tr>
<td>[1, 2, 3, 4, 5]</td>
<td>[!]</td>
</tr>
</tbody>
</table>

- Generates a set of all possible prefixes with a length greater than 1
- Predicts a single suffix event following the prefix.
Four data sets were used to carry out the deep learning approaches

- Three were event logs used for the Business Process Intelligence Challenges (BPIC) from 2012, 2013 and 2014
- Fourth was a help desk event log for an Italian company’s ticket management process
- Event log used for BPI 2012 is an application procedure for financial services
- BPI 2013 used VINST an incident management system for IT related incidents
- BPI 2014 is an event log for different ICT processes (interactions, incidents, changes) for Rabobank Group, a banking and financial services company
## Case Study - Accuracy Results/Evaluation

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>Cases</th>
<th>Events</th>
<th>Prefix</th>
<th>Teacher Forcing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Time (Mins)</td>
<td>Accuracy</td>
</tr>
<tr>
<td>BPI 2012</td>
<td>7469</td>
<td>6</td>
<td>17.9</td>
<td>68.64%</td>
</tr>
<tr>
<td>Help Desk</td>
<td>3803</td>
<td>9</td>
<td>2.7</td>
<td>81.16%</td>
</tr>
<tr>
<td>BPI 2013</td>
<td>7553</td>
<td>13</td>
<td>32.1</td>
<td>65.66%</td>
</tr>
<tr>
<td>BPI 2014</td>
<td>6000</td>
<td>69</td>
<td>42.5</td>
<td>48.28%</td>
</tr>
</tbody>
</table>
Analysis

- The **Prefix** method is shown to have slightly greater accuracy than the **Teacher Forcing** methodology.
- However the **Teacher Forcing** approach takes a far shorter time to train.
- The **Helpdesk** data set has the greatest accuracy due to having the smallest vocabulary of events and set of cases.
- To the best of our knowledge, LSTM networks have not been previously applied to event prediction for the 2014 data set.
Thank you for watching

Stephen Jacob