Large Scale Log Analytics through ELK

Marcel den Reijer

University of Amsterdam
MSc System and Network Engineering

July 5, 2018
Introduction

Problem statement

• Events
• Centralized location
• ELK-stack
  • Elastic Search
  • Logstash
  • Kibana
• Lambda Architecture
  • Batch layer (raw data sets and pre-compute Batch views)
  • Speed layer (Real-time views)
  • Serving Layer (ad-hoc queries)
According the introduction, the following research question is defined as:

- *Which way of applying Artificial Intelligence in the form of Machine Learning/Deep learning can be used to find relevant actionable information from event data suitable for use within the batch layer in a Lambda Architecture?*

This main research question are divided into the following sub questions:

- Which Deep Learning (DL) model fits for this type of data?
- How is the accuracy of the DL model calculated?
- How can the model be tuned during operations?
• Keras Deep Learning API
  • High-level implementation of Artificial Neural Networks (ANN) written in python
  • Run on top of Tensorflow, Theano or CNTK
• Define layers
  • Sequence layer
  • Function layer
• Compile model
  • How to learn - Optimizing Stochastic Gradient Decent algorithms?
    • SGD
    • RMSprop
    • Adagrad
    • Adadelta
    • Adam
Scherer, R. Lstm recurrent neural networks for short text and sentiment classification. 2017

Ruder, S. An overview of gradient descent optimization algorithms. 2018
Technical background

Recurrent Artificial Neural Network and Artificial Neural Network

**Figure:** Recurrent vs Forward neural network; *Source:* [1]
Technical background
Long Short-Term Memory

Figure: RNN sequence; Source: [2]

Figure: LSTM memory cell; Source: [3]
Technical background

Dense layer vs. Dropout layer

Figure: Dense vs Dropout layer; Source: [4]
Technical background
Cross-entropy, Epochs, Batch size and Softmax

- Categorical Cross-entropy
  - Loss Function
  - Indicates distance between what the model the output distribution should be and what the original distribution really is

- Epoch
  - Entire data set went through the algorithm in number of times

- Batch size
  - The amount of test/training examples in one forward or backward pass
  - RAM memory

- Softmax function
  - Calculate the probability distribution of each target class over $n$ different/possible target classes for the given inputs
Proposed implementation

PoC

**Figure: Architecture**
Proposed implementation

PoC layers

Figure: AI layers

- Advantages of this model
  - Interfaces may be changed easily
  - Change AI model easily
  - Change Algorithms easily
Results

20 Epochs

Accuracy per Epoch

Loss per Epoch
Results

40 Epochs

Accuracy per Epoch

Loss per Epoch
### Results

#### Table: Total accuracy and loss per model as a whole

<table>
<thead>
<tr>
<th>Model</th>
<th>Loss</th>
<th>Accuracy</th>
<th>Loss</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>2.96977%</td>
<td>16.36363%</td>
<td>2.21985%</td>
<td>29.81818%</td>
</tr>
<tr>
<td>SGD</td>
<td>3.85677%</td>
<td>10.39999%</td>
<td>3.75910%</td>
<td>10.39999%</td>
</tr>
<tr>
<td>RMSprop</td>
<td>3.00070%</td>
<td>5.127276%</td>
<td>2.73776%</td>
<td>19.41818%</td>
</tr>
<tr>
<td>Adadelta</td>
<td>3.44532%</td>
<td>10.54545%</td>
<td>3.11788%</td>
<td>13.81818%</td>
</tr>
<tr>
<td>Adagrad</td>
<td>3.52502%</td>
<td>10.39999%</td>
<td>2.28705%</td>
<td>27.27272%</td>
</tr>
</tbody>
</table>
• Discussion
  • Adam is the best choice according the results
  • Does not guarantee the correctness.
  • Model needs a lot of data and time in order to get a better accuracy and learning rate

• Conclusion
  • Apply LSTSM with Adam as optimizer
  • Softmax
  • Inject false or true events, in such a way the accuracy will change
Future work

- Apply AI on Speed layer
- Optimize AI module
- Investigate better approach for accuracy and predictions.
- What should be the time-period to analyze to optimize relevant context but also trying to be as near-realtime as possible?
Questions?
De Mulder, W., Bethard, S. & Moens, M.F.  
Deep learning applications and challenges in big data analytics.  

C. Godbout.  
Recurrent Neural Networks for Beginners.  

A. Hassan, A. & Mahmood.  
Deep learning for sentence classification.  

Dropout: A simple way to prevent neural networks from overfitting.  