Twitter Sentiment Analyzer with Machine Learning

Background

Obtaining feedback from a few of your customers is easy. However, imagine that you could ask the whole world! You would know immediately the general opinion and could react accordingly. This machine learning solution examines twitter data in real-time for their tonality. Combined with a query for your product or service this yields an up-to-the-minute dashboard of your customers' satisfaction. The classification, whether a tweet is positive or negative, is done by a deep neural network.

The idea to train a deep neural network already exists since 1990. As machine learning technology progresses, this idea transforms into reality, what is known as deep learning today.

Dataset

• Tweet-sentiment140-dataset of Stanford University
• 1.6 million preclassified tweets

Clean & format data

• Each tweet has positive or negative polarity
• Lexicon-based numerical/numerical vector transformation
• Example to create a vector (Bag-of-Words-Model):

<table>
<thead>
<tr>
<th>Lexicon</th>
<th>Tweet</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>dog, cat, house, flower</td>
<td>The hungry dog runs into the house.</td>
<td>[1,0,1,0]</td>
</tr>
</tbody>
</table>

Define neural network

• Deep learning framework TensorFlow
• Feedforward network
• One hidden layer
• Input layer: 2638 neurons = length of lexicon
• Output/softmax layer: 2 neurons (positive or negative)

Optimize the training

• L2 Regularizer
• Dropout: Dropping out units in the neural network to reduce overfitting

Use model

The training process of an ML model uses the training data to learn from. The model is an artifact and an output of this process. For using the trained model the weights and parameter will be restored.

Generate predictions

In this application a real-time prediction is generated, so the front-end can use the results interactively. Another option would be a batch prediction, when you want to obtain predictions for the observations all at once.

To feed the application synchronously with real data the pipeline has to be connected to a streaming API or stream processing framework.

Sentiment Analysis

• Type of data mining that measures the inclination of people's opinions
• Quantify the general public's sentiments in Twitter
• Typical use cases:

  - Read the last classified tweet out of the database
  - Process the user input
  - create statistics and diagrams

Visualize via web-front-end

• Real-time pie chart
• Real-time sentiment line
• Last five tweets
• Matches for the keyword

Present the results

Visit our website: https://novatec.sentimentanalyzer.de