Sentimental Smacking:
An ML Toolkit Exploration

Jarred Bettencourt & Jeffrey Tran

Problem: Neural Net Model Visualization

- Sentiment analysis: what is it?

- Non-programmers ➜ Are Neural Net visualization toolkits useful for beginners?

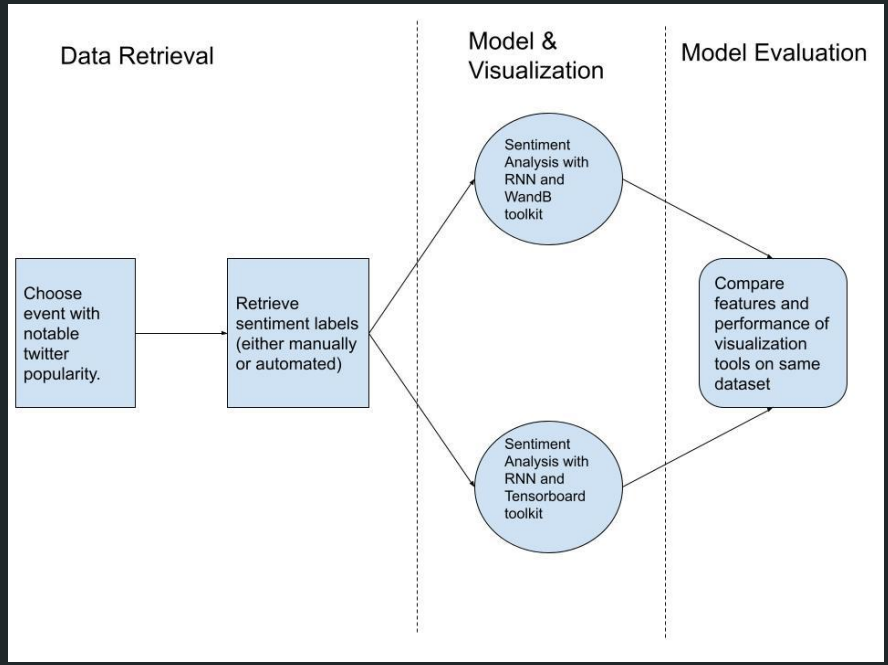
- How do we answer this question?
  - ➜ Create a model on data from particular event, and compare toolkits

What event was that?

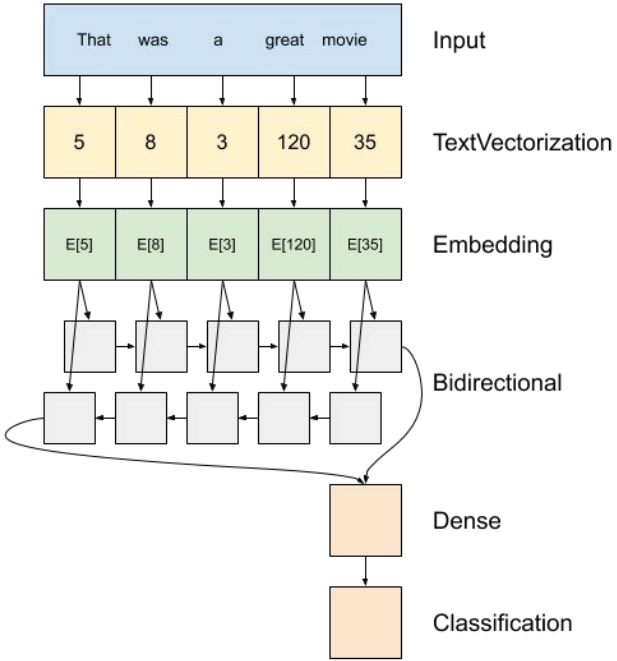
Pictured: Will Smith striking Chris Rock after making a joke about Jada Pinkett Smith’s lack of hair due to alopecia at the 2022 Oscars ceremony.

Approach

- Data Retrieval ➜ Manually labeled tweets from twitter
- Model ➜ Recurrent Neural Network adapted for sentiment analysis
- Visualization ➜ Results of RNN were visualized with WandB and Tensorboard
- Evaluation ➜ Performance and features of visualization tools were compared



S.A. Recurrent Neural Net



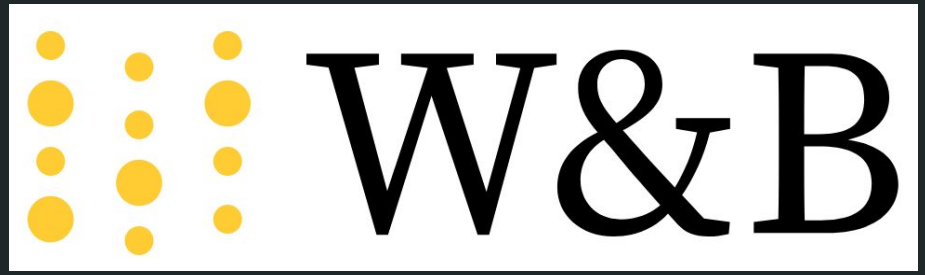
- Recurrent Neural Network (RNN) as template for sentiment analysis
- RNNs are very well suited for temporal data (most notably sentences!)

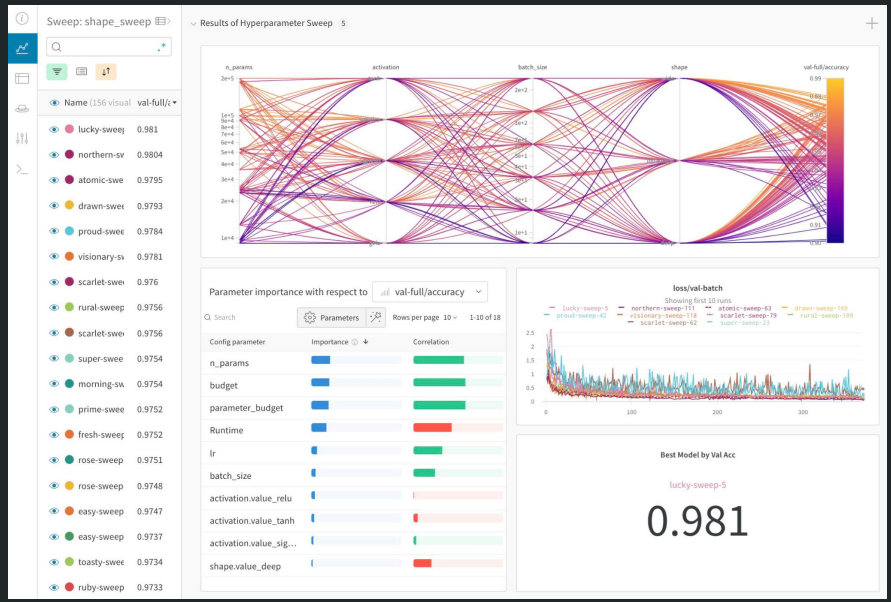
Steps:

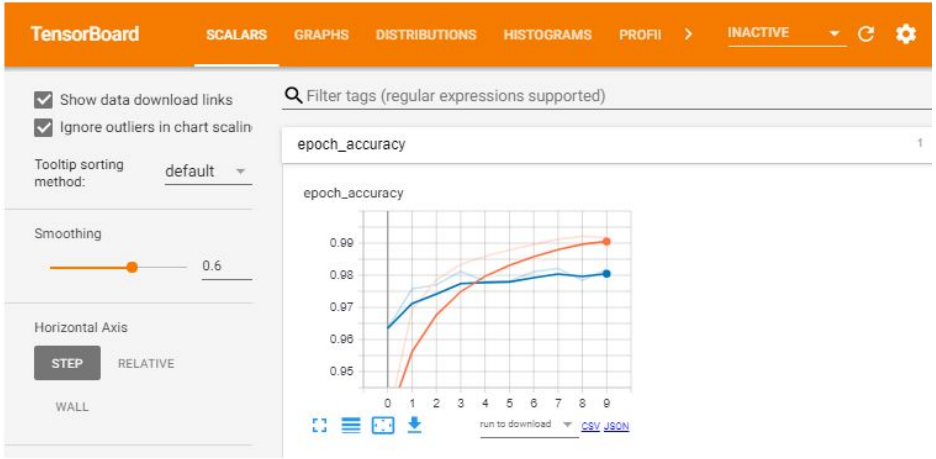
1. Create Vocabulary from tweets
2. Feed sentences into Neural Net
    a. Embedding Layer
    b. Bidirectional Layer
    c. ReLU Activation Layer
    d. Classification Layernt
3. Receive Classification

Tool #1: Weights and Biases

- Open source deep learning visualization toolkit
- Provides large degree of control over many deep learning tasks, including
  - Hyperparameter Optimization
  - Experiment Analysis
  - Collaborative Reports
- Includes system analytics
- Can be self hosted or cloud based





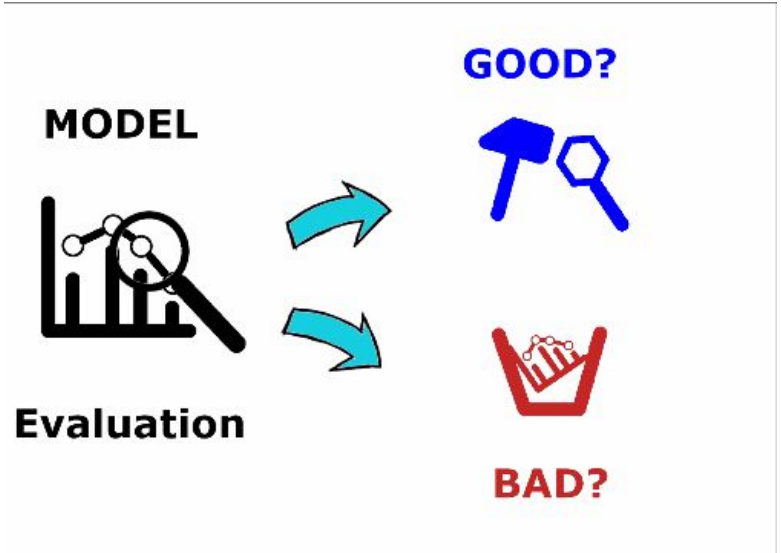


Tool #2: TensorBoard

- Included with popular deep learning library TensorFlow
- Shows similar plots and visualizations to WandB, but must be tailored
- No semblance of system analytics
- Must be locally hosted
- Reliant on old package versions

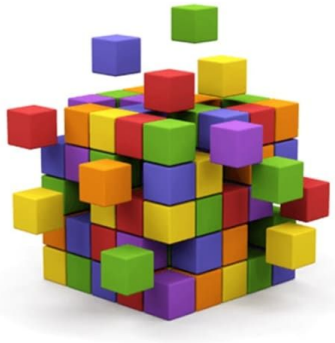
Model Evaluation (Neural Net)

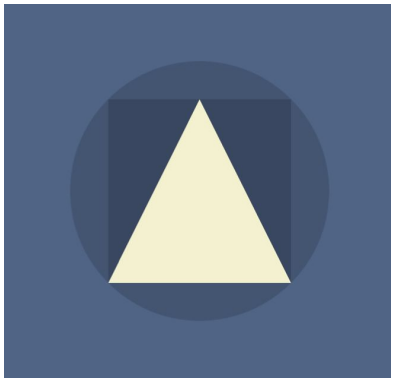
$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$



- After 150 epochs with a vocab size of 50, S.A. RNN classified tweet sentiment with around 87%+ accuracy
- Could achieve higher accuracy with higher vocab size → Overfitting!
- RNN optimized on Binary Cross Entropy
- Evaluation on training set

Software Metrics

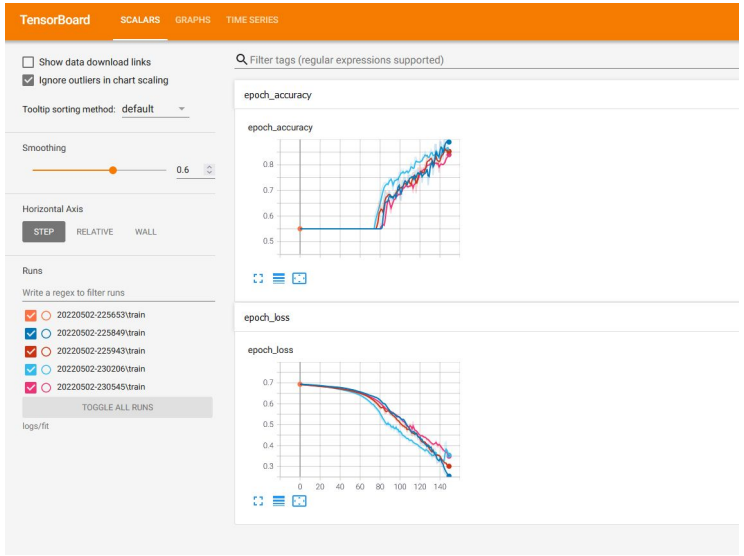


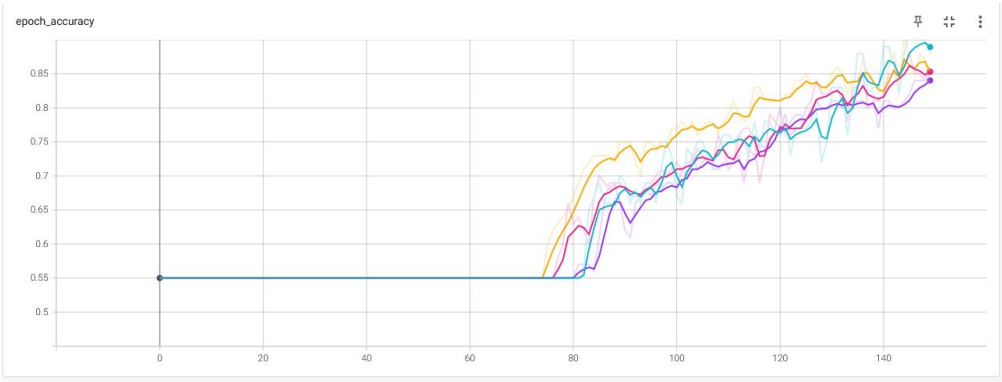


Ease of Use - Extensibility - Usability

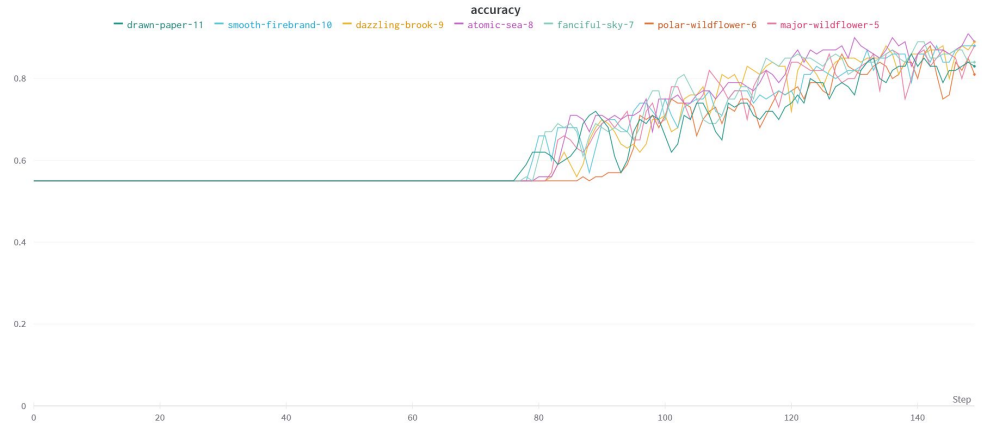
TensorBoard Evaluation

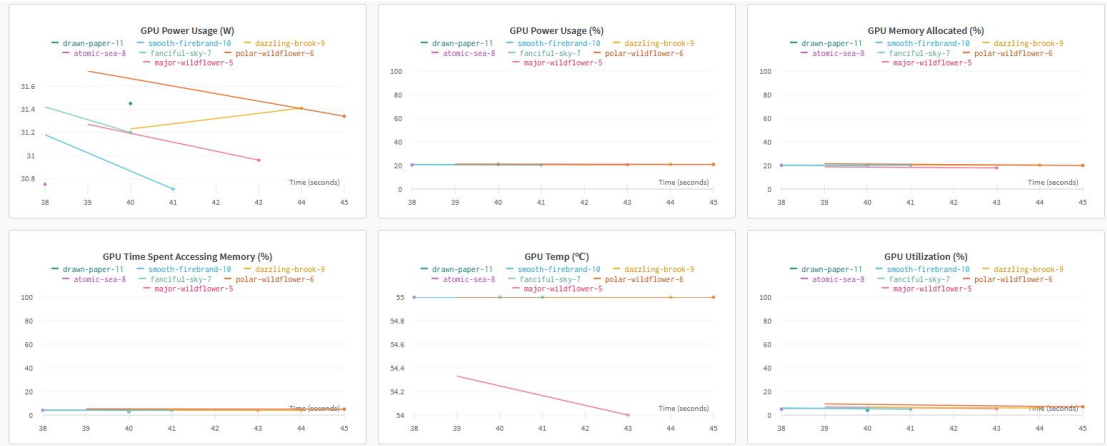
- Ease of Use (5/10):
  - Requires downgrading Google-auth
  - Large amount of customization needed
- Extensibility (7/10):
  - Large amount of custom plots
  - Only works with TensorFlow
- Usability (3/10):
  - Only analyzes log files
  - No real interaction with model
- Overview: Not great for non-ML people!





Weights and Biases Evaluation





- Ease of Use (10/10)
  - Nearly plug and play
  - Visualization templates available
- Extensibility (8/10):
  - Indefinite amount of customization
  - Works with any deep learning library
- Usability (9/10):
  - System and Model analytics
  - Can interact with models from WandB GUI
- Overview: Amazing learning tool and visualizer/optimizer for any level!

Discussion

Data:

- Relevance waned very quickly (3 days)
  - High tweet engagement → Chris Rock
  - Low tweet engagement → Will Smith

Visualization kits:

- Almost no use case where WandB is not superior to TensorBoard

- Universally true for programmers and non-programmers

Thank you for listening!