Categorizing Tweets

Estella Dentinger, Nikita Singh, Andrew Hu, Brittany Nguyen

Time Magazine's
2021 Person of the Year

Time Magazine's 2021 Person of the Year

our favorite "Transformer"

– 2:30pm on Sep 19, 2012

Elon Musk

@elonmusk · **Follow**

7 engine static fire

– 2:30pm on Sep 19, 2012

7 engine static fire

– 2:30pm on Sep 19, 2012

What was he tweeting about?

7 engine static fire

– 2:30pm on Sep 19, 2012

This “fire” was not describing an accident

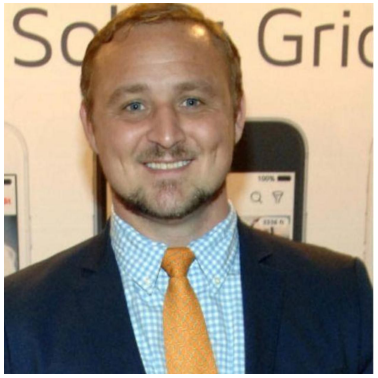
But many are tweeting about
real live emergencies right now

Body of missing CEO found over a year after he texted “911” and vanished

– May 9, 2023

The remains were found in the 2900 block of Santa Monica Boulevard, the area's police department said on Facebook. The body was identified as Mann's on May 6.

In December 2021, Mann was reported missing to the Los Angeles Police Department, officials said. The CEO was last seen on Nov. 30, 2021. According to CBS Los Angeles, Mann was last seen outside a convenience store in Studio City, wearing a blue baseball cap, dark sweatshirt and pants and black shoes. According to CBS Los Angeles, Mann texted "911" shortly after leaving the store.



Source: Breen, K. 2023. Body of missing CEO found over a year after he texted "911" and vanished. CBS News. https://www.cbsnews.com/news/body-of-beau-mann-missing-ceo-sober-grid-found/?ftag=CNM-00-10aab8d&linkId=213726302.

Business Use – Time Matters

Reduce Property Damage

For fire emergencies, **each minute** of faster response time = **$5,000 to $10,000** reduction in property damage[1]

Only **58 % of fire departments** in the U.S. meet NFPA's **six-minute response** guideline, from alert to arrival[1]

Save Priceless Lives

U.S. statistics each year[2]

- **~50 million** medical assistance calls
- **~12 million** life-threatening calls

European Resuscitation Council on cardiac arrest: "defibrillation within **3–5 min of collapse** can produce survival rates as high as **50–70%**"

Transmit Mobile Location

911 operators **cannot automatically determine** the location of your mobile call[3]

In Europe, regulation only started requiring **new phones** to have the technology recently, starting in March 2022[4]

Many Tweets and social media posts include GPS location automatically

Source: [1] Purvis. Current State of Turnout Times. [2] RapidSOS. 2015. Outcomes Quantifying the Impact of Emergency Response Times. [3] Wikipedia. 2023 May 08. Enhanced 9-1-1. [4] ETSI. 2022 March 22. Advanced Mobile Standard.

Technical Challenges

Meaning & Position Embeddings Not Enough

- Consider: “The best **movie is Ablaze**!” vs. “The theater showing the **movie is ablaze**!”
- “Ablaze” present in both tweets, in similar positions

N-grams Inadequate

- N-gram “movie is ablaze” is present in both tweets
- Testing showed no improvement in out-of-sample performance by adding bigram preprocessing

Limited by 140 or 280 Characters

- Tweets often include abbreviations, misspellings, or out of context phrases; **hard to capture true context in few words**
- Transformers pre-trained on prose, and no “translator” on the internet for: Tweets <> Sentences

Our Modeling Choices

Methods

- Bag-of-Words
- GloVe Embedding
- HODL Transformer
- BERT Transfer Learning
- Stacked BERT and Bag-of-Words

Data Description

- Kaggle: ~10,000 tweets, pre-labeled, with common emergency phrases for both true and false positives
- No Information Rate = 57% (split of 0s and 1s)
- Columns: tweet text, location sent from (if available), binary output variable, keywords associated with emergency in the tweet (not used in best model)

Source: Howard, A., devrishi, Culliton, P., Guo, Y. 2019. Natural Language Processing with Disaster Tweets. Kaggle. https://kaggle.com/competitions/nlp-getting-started.

1. Bag-of-Words

	Model 1: Unigrams	Model 2: Bigram Text + Keywords
Training Accuracy	0.96	0.95
Validation Accuracy	0.79	0.76

- Experimenting with Context:
  - Model 1: Unigram STI tokens only
  - Model 2: Bigrams STI + keywords input
- Model Setup:
  - Dense layer with 8 hidden units and ReLU activation
  - Output layer with sigmoid activation function
  - 40,026 trainable parameters
- Hyperparameter Tuning:
  - 16 hidden units
  - 20 epochs

2. GloVe Embeddings

I can capture complex semantic relationships!

Just needed a quick summary...

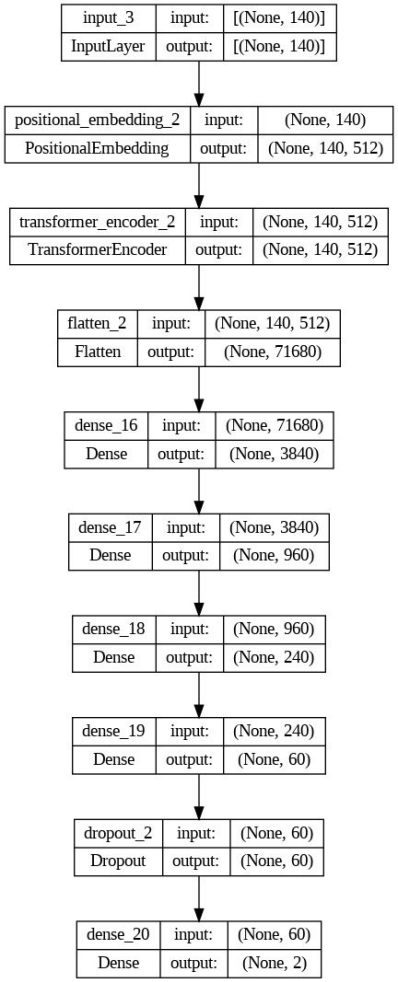
	DNN with GloVe Embeddings
Training Accuracy	0.94
Validation Accuracy	**0.78**

- **Model Setup:**
  - **Pooling layer**
  - **Dense layer with 8 hidden units and ReLU activation**
  - **Output layer with sigmoid activation function**
  - **500,826 trainable parameters**
  - **Dropout layer regularization**

Source: Image from Dall-E. Joke from GPT-4.

3. HODL Transformer

	HODL Transformer Performance
Training Accuracy	0.57
Validation Accuracy	0.57



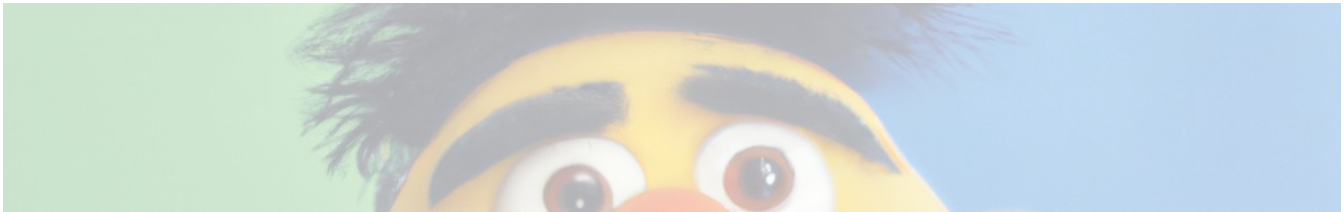
- Model Setup:
  - Max_text_length of 140 (tweets limited to 140 characters)
  - 5 self-attention heads
  - Four dense layers with ReLU: 3840, 960, 240, 60 hidden units
  - Dropout layer regularization
  - Softmax output layer with 2 categories
  - 293,111,014 trainable parameters
  - Batch size of 128 for 10 epochs

- Training ended up applying naive rule = predict all 0s (majority class) = No Information Rate = 0.57
- Model did not find a signal, with limited dataset + no pre-trained embeddings

4. BERT Transfer Learning

	Model 1: BERT Transfer Learning with 1 Dense Layer	Model 2: BERT Transfer Learning with 2 Dense Layers
Training Accuracy	0.83	0.84
Validation Accuracy	**0.816**	**0.814**

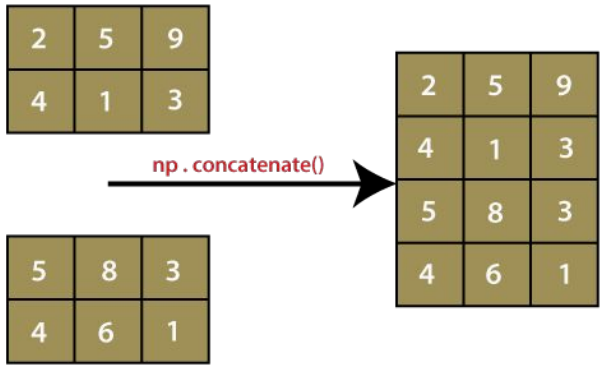
- Model Setup:
  - 2 hidden layers with 64 units each
  - Dropout layer regularization
  - Output layer with sigmoid activation function
  - Batch size of 32 for 20 epochs
  - 53,506 trainable parameters
- Test Set Accuracy: 0.809



5. Stacked BERT and Bag-of-Words

	BERT Transfer Learning with Bag-of-Words Performance
Training Accuracy	0.96
Validation Accuracy	**0.75**

- **Model Setup:**
  - **Concatenate outputs of both models, and feed it through a dense layer**
  - **Dropout layer regularization**
  - **Dense layer with 32 neurons**
  - **Output layer with sigmoid activation function**



Source: Image from https://www.javatpoint.com/numpy-concatenate.

Live Demonstration

Key Learnings

More complexity & parameters ≠ Improved performance
Not a linear relationship (Transformer/Stacked < BERT TL 81%)

Start with Pre-trained Embeddings & Transfer Learning
when data is limited (tweets/abbreviations)

Tuning is an art, requires experimentation (and A100 GPUs)
Future steps: fine tuning with Davinci

Thank you

"Next time, don't forget the dropout layer."

Cartoon by Robert Leighton, The New Yorker.