# Fortune Cookie Text Generation

Dora Dong

Brianna Brown

Thomas George

Tony Tao

# About Us

- Deep Learning Group (Jan, 2019 – Dec, 2019)
  - Weekly events to learn DL techniques and discuss applications.
  - Small projects to enhance our understanding of the algorithms.
- Team
  - Dora Dong,  Data Scientist @Fred Hutch
  - Brianna Brown, Salesforce Administrator @Optimum Energy
  - Thomas George , Software Developer @Optimum Energy
  - Tony Tao, Technical Product Manager @Accolade, Inc.

Dream lofty dreams and as you dream so….

# Our Fortune Cookie Project Goals

- Learn some things about natural language processing

- Get some hands on experience working in a group environment on a machine learning project

- Generate some rad fortunes

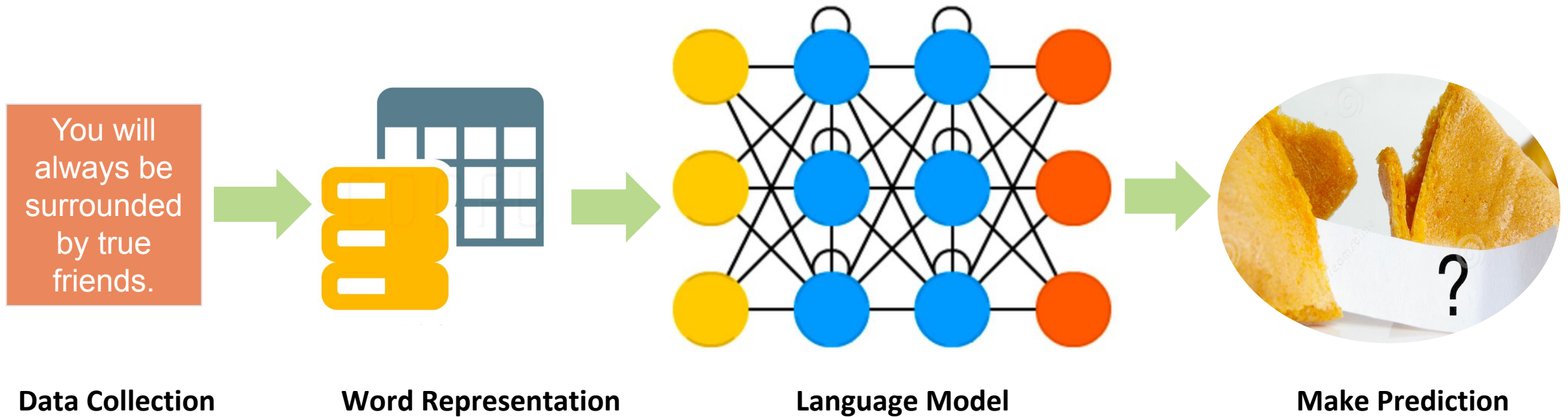Be calm when confronting an emergency crisis

# What we did:

We tried 3 different text generation models with varying degrees of success

1. **Baseline model:** GRU RNN with Embedding Layer trained on training data

2. **GloVe model:** GRU RNN with using a GloVe Embedding Layer

3. **GPT-2 model:** Open AI released project

Who will always be closed again even after

# Overview



You will always be surrounded by true friends.

**Data Collection**　　　**Word Representation**　　　**Language Model**　　　**Make Prediction**

Eyes open up your work interests can capture

# Word representation

x:  "you will be successful in your work".

$$x^{<1>} \quad x^{<2>} \quad x^{<3>} \quad \cdots \quad x^{<7>}$$

| Vocabulary | you | will | be | | work |
|---|---|---|---|---|---|
| $a$ | 0 | 0 | 0 | | 0 |
| $arron$ | 0 | 0 | 0 | | 0 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ |
| $be$ | 0 | 0 | 1 | | 0 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\cdots$ | $\vdots$ |
| $in$ | 0 | 0 | 0 | | 0 |
| $\vdots$ | $\vdots$ | 1 | $\vdots$ | | $\vdots$ |
| $work$ | 0 | $\vdots$ | 0 | | 1 |
| $\vdots$ | $\vdots$ | 0 | $\vdots$ | | $\vdots$ |
| $you$ | 1 | 0 | 0 | | 0 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ |
| $zulu$ | 0 | 0 | 0 | | 0 |

Eyes love you will have a bright future

# The GRU model

**GRU**

reset gate

update gate

sigmoid        tanh        pointwise multiplication        pointwise addition        vector concatenation
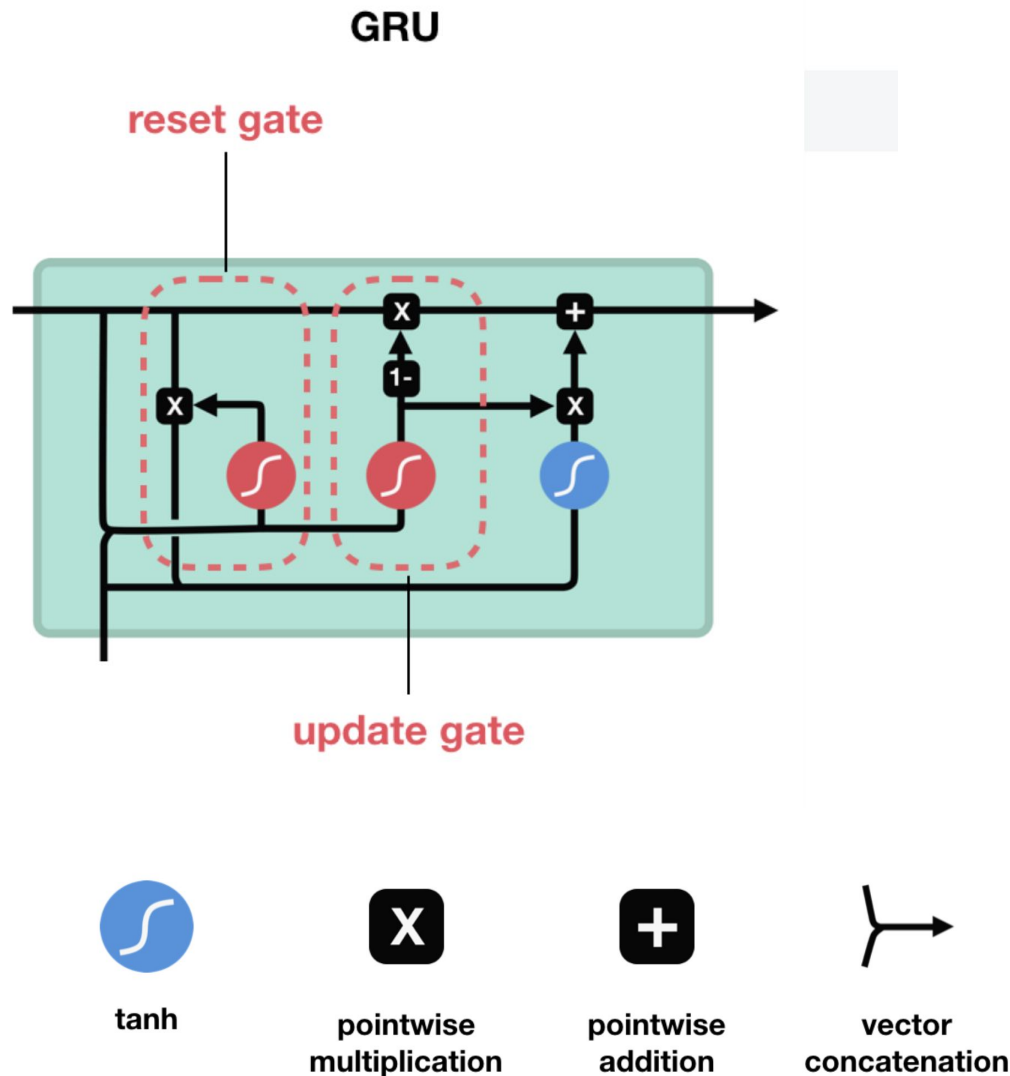
Our models used a memory state to track which words were important for predicting the next word.

# Model Comparison

| Models | Embedding Layer | Output | Thoughts |
|---|---|---|---|
| Baseline Model | Based on corpus (2087 unique words) | Fortune Smiles Upon You Today To Be Tomorrow | 1. No stopping point. 2. Not a complete sentence. |
| GloVe Embedding Layer | GloVe 6B tokens, 400K vocab | Be Calm When Confronting An Emergency To Be | |
| GPT2 Embedding Layer | GPT2 8 million web pages | You seek to understand those who are turning you away, you will be successful in your work, … | 1. It learns the punctuation. 2. Still requires a random prefix, and predefined length. |

# Challenges

- Really small data set (~2,000 sentences)
  - With GloVe and Baseline models it was a challenge to get our model to come up with new fortunes rather than regurgitating existings ones
- It is really difficult to determine metrics for how "fortuney" our generated fortunes were
  - Difficult to determine how well our model is performing
  - Would require human interaction to measure performance
  - There is inherent human bias in that fortune cookies can be both positive or negative
- We were unable to have our model determine a stopping point by itself.

# Takeaways

- Through meeting in this study group we gained a more grounded understanding of the math that drives deep learning

- While our class focused on data modeling, much of our time was spent on data engineering

- We learned that AI projects is a ongoing cycle of failing and trying again

- The black box of deep learning is real. We found some success from using open source tools but our results were not production ready

# And always remember…

# THANK YOU!

# Contact Info

- Contributions
  - Dora Dong, doradong7@gmail.com
  - Brianna Brown, Brianna.t.brown@gmail.com
  - Thomas George, jorge0136@gmail.com
  - Tony Tao, tonyt.00@outlook.com
- Project Link:
  - https://github.com/WomenInDataScience-Seattle/FortuneCookie
- Data Circle Webpage
  - www.datacircle.org