Product Review Summarization

By: Allen Ye and Andrew Miranda

Problem

- Reading reviews = lots of time, not efficient
- Information overload when making purchase decisions using product reviews

Motivation

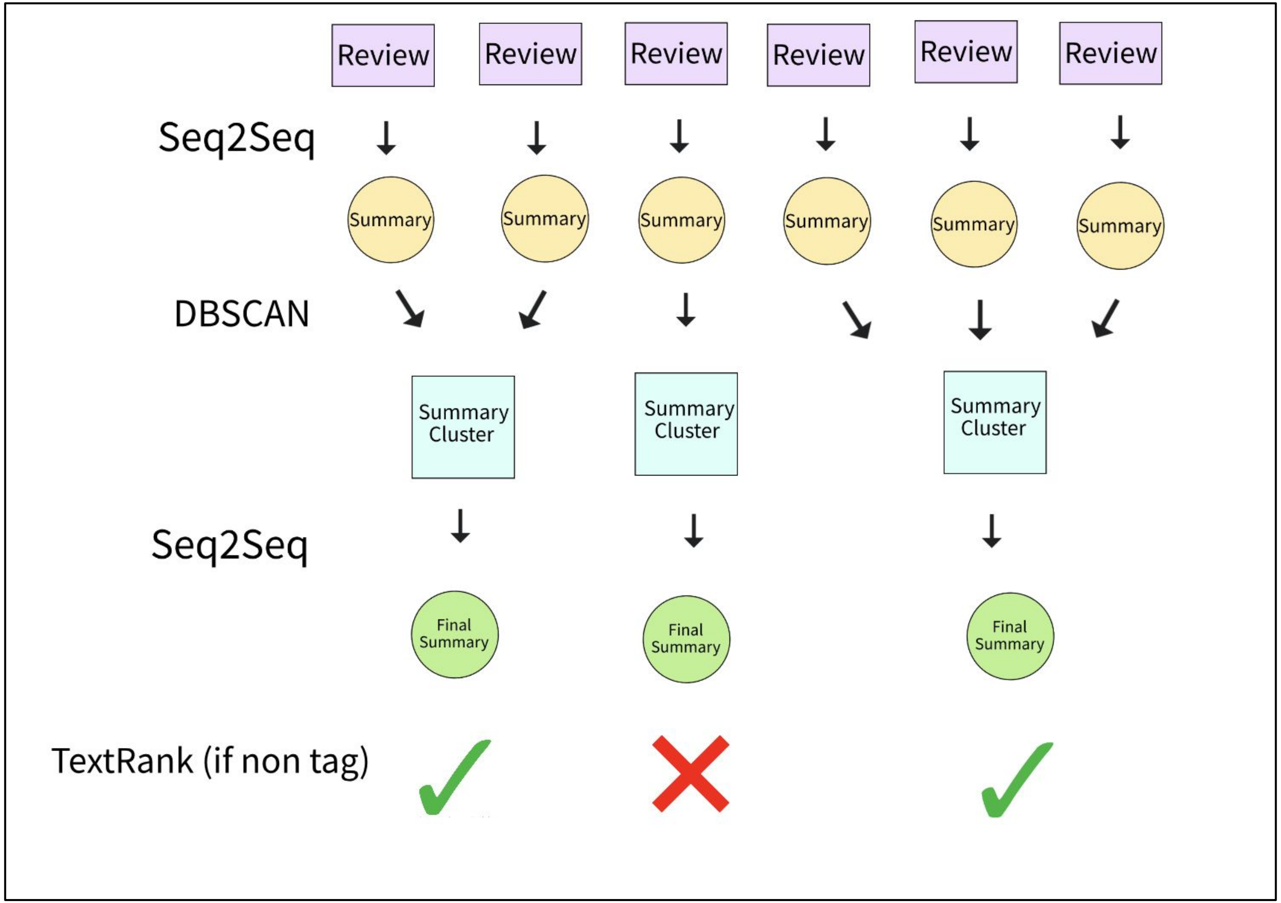
- Develop an application that can automatically summarize product reviews
- Save people's time by providing concise summaries, extracting key sentiments, and generating product-specific tags

Our Approach

- 2 types of summaries for a product: sentences and tags
    - sentences: a few short sentences describing the product
    - tags: important aspects of the product
- No available dataset for this goal, creative approach:

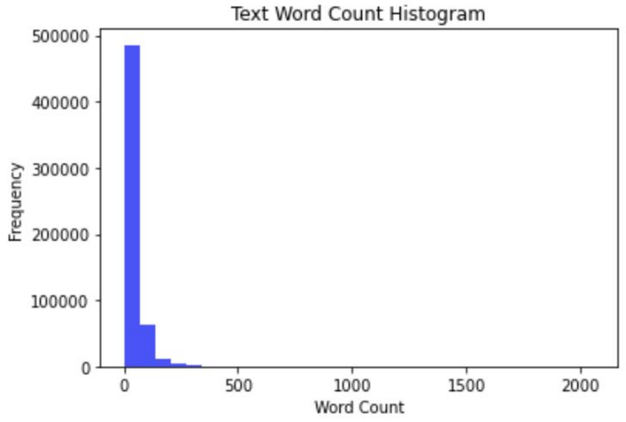
1. Lots of reviews
2. Summarize each review
3. Group/cluster summaries
4. Re-summarize clustered summaries

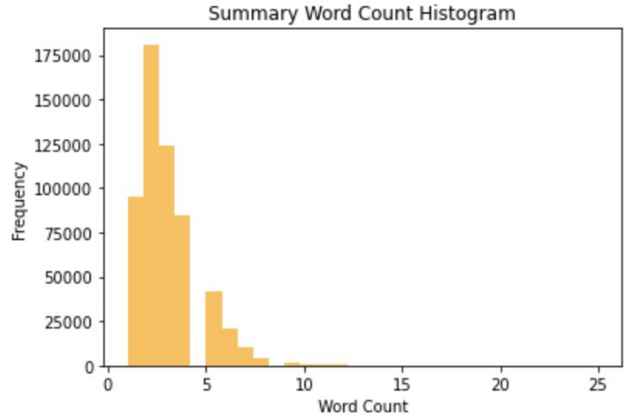
E2E flow diagram



Data + Preprocessing

- Dataset: Amazon Fine Food Reviews
    - Mainly food items
    - ~500,000 rows of reviews + metadata
- Preprocessing:
    - removing stop words + punctuation
    - lowercasing
    - adding tokens (SOS, EOS)
- Filtering:
    - max text length = 150 tokens
    - max summary len=12 tokens
- Dataset Subsetting:
    - two different output types = train on different data
    - sentence model: min summary len>= 3 tokens
    - tag model: min summary len <= 3 tokens

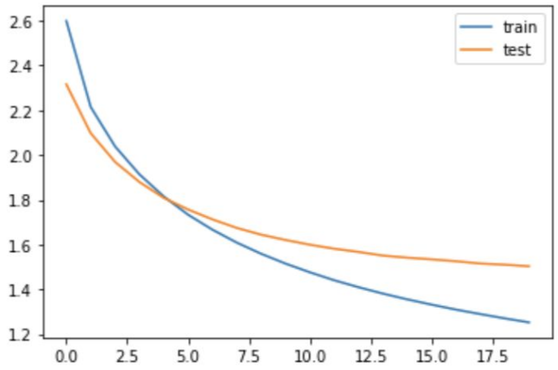




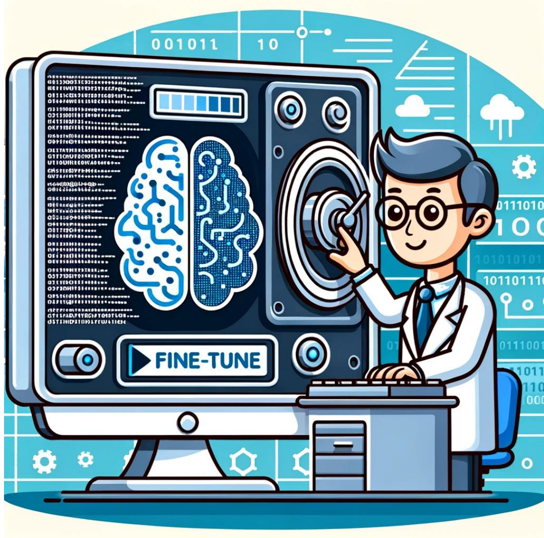
Model + Training

- For tokenizers, excluded rare words
- Seq2Seq model
  - LSTM based encoder-decoder architecture
  - attention layer
  - optimizer: rprop
  - loss: sparse categorical cross entropy
- Total params: ~19 million
- Inference:
  - Decoding Process: Utilizes the trained model to decode each review into a summary, word by word.
  - Attention Mechanism: Focuses on relevant parts of each review for meaningful summarization.

```
Model: "model"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
input_1 (InputLayer)            [(None, 150)]        0
__________________________________________________________________________________________________
embedding (Embedding)           (None, 150, 100)     5815400     input_1[0][0]
__________________________________________________________________________________________________
lstm (LSTM)                     [(None, 150, 300), ( 481200      embedding[0][0]
__________________________________________________________________________________________________
lstm_1 (LSTM)                   [(None, 150, 300), ( 721200      lstm[0][0]
__________________________________________________________________________________________________
input_2 (InputLayer)            [(None, None)]       0
__________________________________________________________________________________________________
lstm_2 (LSTM)                   [(None, 150, 300), ( 721200      lstm_1[0][0]
__________________________________________________________________________________________________
embedding_1 (Embedding)         (None, None, 100)    1397900     input_2[0][0]
__________________________________________________________________________________________________
lstm_3 (LSTM)                   [(None, 150, 300), ( 721200      lstm_2[0][0]
__________________________________________________________________________________________________
lstm_4 (LSTM)                   [(None, None, 300),  481200      embedding_1[0][0]
                                                                 lstm_3[0][1]
                                                                 lstm_3[0][2]
__________________________________________________________________________________________________
attention_layer (AttentionLayer ((None, None, 300),  180300      lstm_3[0][0]
                                                                 lstm_4[0][0]
__________________________________________________________________________________________________
concat_layer (Concatenate)      (None, None, 600)    0           lstm_4[0][0]
                                                                 attention_layer[0][0]
__________________________________________________________________________________________________
time_distributed (TimeDistribut (None, None, 13979)  8401379     concat_layer[0][0]
==================================================================================================
Total params: 18,920,979
Trainable params: 18,920,979
Non-trainable params: 0
__________________________________________________________________________________________________
```



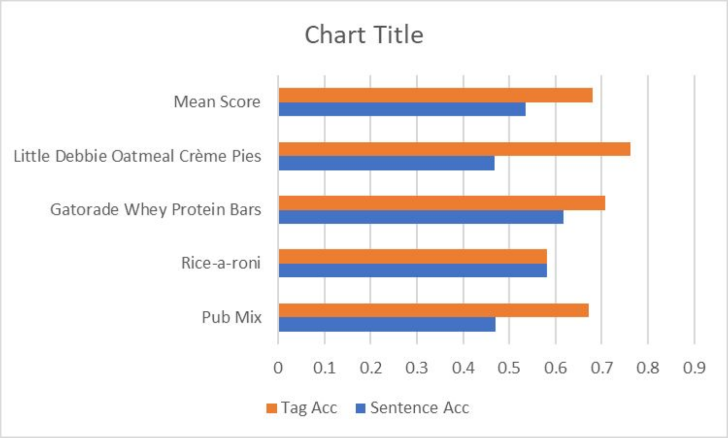
Fine Tuning



- different amount of LSTM layers
- different embedding dimensions
  - higher = better, but limited by Kaggle's memory + GPU quota
- smaller batch sizes
- different dropout values:
  - too low = model more likely to copy/reuse what original label is
  - too high = model more likely to output very broad summaries (good product)
  - had to find a good balance (0.2-0.3 range)
- Training two different models
  - one on shorter summaries for tags, one on longer summaries for sentences

Results + Evaluation

- we use word embedding similarity as our metric
  - effective metric for comparing summaries and tags because it captures semantic relationships between words
- for tag output, our mean score is **0.681**
- for sentence output, our mean score is **0.534**
- drawbacks:
  - not enough reviews (we ideally need ALL of the reviews for the product we are evaluating)
  - not enough different products (bottleneck was time since we had to manually collect)



Example Output

2 VIDEOS

Roll over image to zoom in

Little Debbie Oatmeal Creme Pies (24 Count of 2.6 oz Cookies) 62.4 oz

Visit the Little Debbie Store

4.5 ★★★★☆ 2,083 ratings | 30 answered questions

Flavor Name: **Oatmeal**

Oatmeal | oatmeal creme | Chocolate Chip

Size: **2.6 Ounce (Pack of 24)**

1.01 Pound (Pack of 1) | 1.01 Pound (Pack of 6)
1.01 Pound (Pack of 16) | 1.98 Pound (Pack of 2)
1.99 Pound (Pack of 3) | 2.6 Ounce (Pack of 24)
9.7 Ounce (Pack of 1)

Occasion	Wedding, Anniversary, Birthday
Flavor	Oatmeal
Number of Items	1
Unit Count	62.4 Ounce
Brand	Little Debbie

About this item

- Two soft, chewy oatmeal cookies baked with whole grain oats and molasses, then layered with creme.
- Each delicious oatmeal creme sandwich comes individually wrapped for freshness, and packaged in a box of 24 packs.
- For a unique twist, these wholesome snacks can be used in recipes, added as a special topping, or warmed up in the oven.
- A great snack to pack with your lunch or enjoy on the go at any time of the day.
- The very first snack made by Little Debbie and proudly baked in the USA since 1960.

Amazon summary:

Customers say

Customers like the taste and quality of the snack cake. They mention that it's a great snack and not overly sweet. They appreciate the softness and delivery. However, some customers have different opinions on value, size, and freshness.

AI-generated from the text of customer reviews

Taste | Quality | Delivery | Softness | Freshness | Value | Size | Hardness

Our sentence summary:

```
'Healthy tasty snack. Best cookie ever.'
```

Our tag summary:

```
['great taste', 'good value', 'great product',
'good', 'best', 'great value']
```

Future Work

- larger evaluation dataset
- different dataset
- apply to other types of reviews
    - service
    - restaurant
    - location

Thanks for listening!