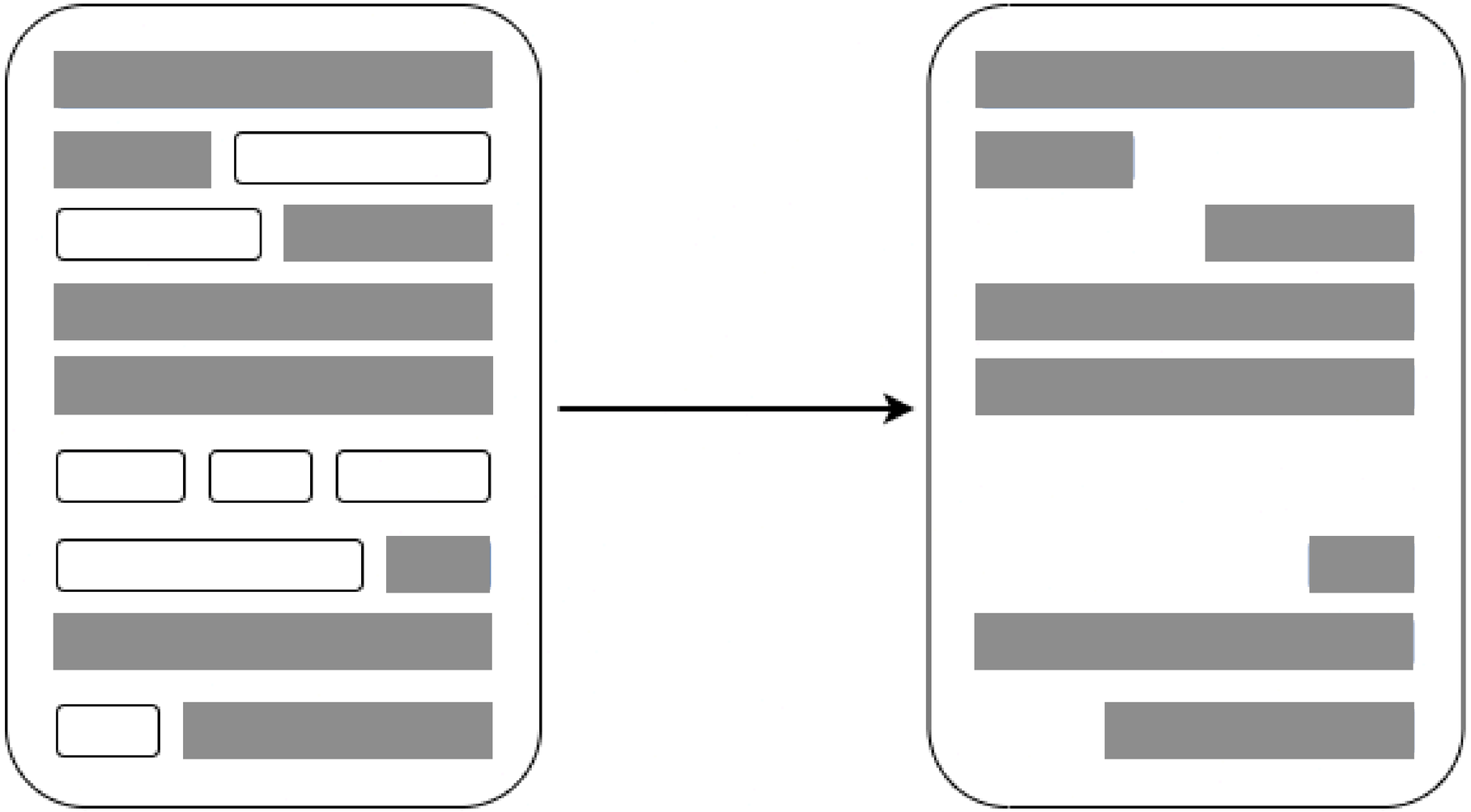
SUMMARUNNER

Extractive Summarization with Recurrent Neural Networks



PROBLEM STATEMENT

- With a lot of information on the web, the need for document summarization is a must and increasing. Summarization models help extract the most important information from a long document and present it concisely.

- Our project focuses on extractive summarization, which involves selecting sentences directly from the document that are most relevant to the summary.

DATASET OVERVIEW

CNN/DailyMail

Features:

Articles - Original Text

Highlights - Human Written Summaries

ID - Unique Identifier

Why This Dataset?

Widely used

More than 250K articles

Long form articles - more suitable for summarization

MODEL EXPLANATION

Word Level RNN

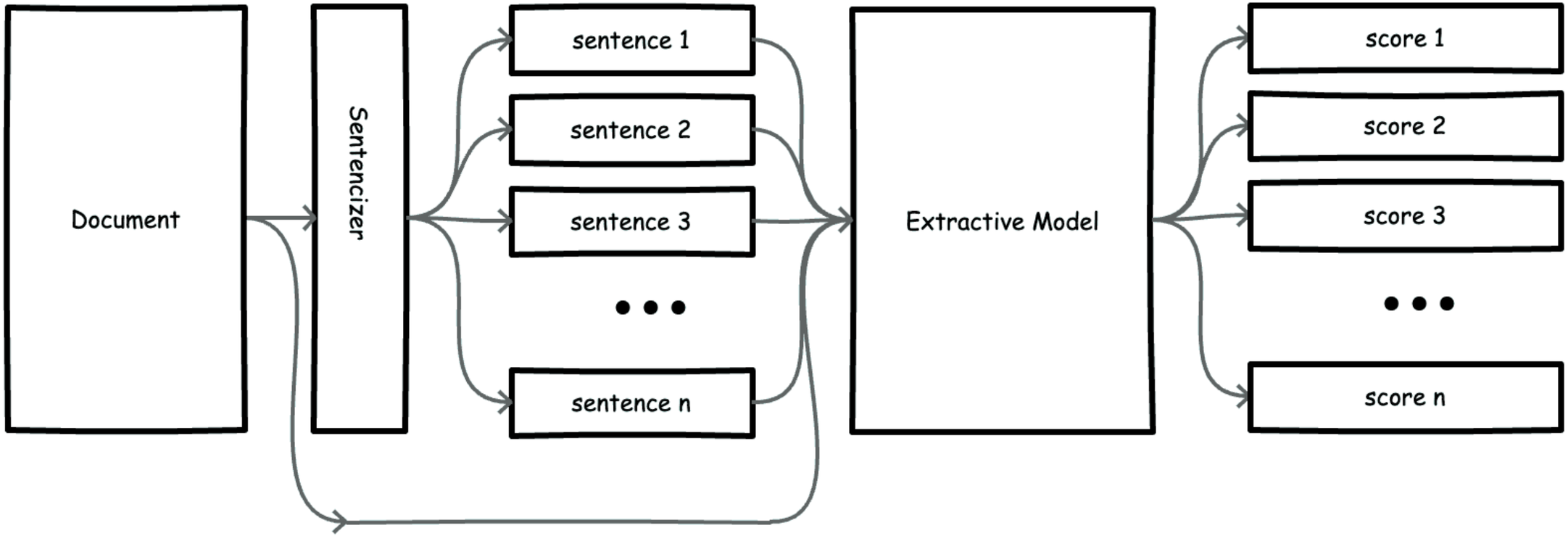
- Bidirectional GRU to process word embeddings
- Output: Sentence level representation

Sentence Level RNN

- Another Bidirectional GRU to process sentence representation
- Output: Probability for each sentence

Linear Layer

- Produces the final decision whether to include the sentence in the summary or not


Document
Sentencizer
sentence 1
sentence 2
sentence 3
sentence n
Extractive Model
score 1
score 2
score 3
score n

TRAINING AND LOSS FUNCTION

- **Training:**
  - The model is trained using binary cross-entropy loss, which is suited for binary classification tasks (sentence inclusion vs. exclusion).
  - Adam optimizer is used for efficient training, adjusting model weights during backpropagation.

- **Loss Function:**
  - **BCEWithLogitsLoss:** This loss function is used because the task involves binary classification (whether to include each sentence in the summary or not). It combines a sigmoid activation and binary cross-entropy in a single function.

EVALUATION METRICS

- **ROUGE** (Recall-Oriented Understudy for Gisting Evaluation)**:**
- **ROUGE** is the most widely used evaluation metric for summarization tasks. It compares the n-grams (uni-grams, bi-grams, etc.) and sequences between the generated and reference summaries.
- **ROUGE-N:** Measures the overlap of n-grams between the predicted and reference summaries.
    - **ROUGE-1:** Measures uni-grams (single words).
    - **ROUGE-2:** Measures bi-grams (two consecutive words).
- **ROUGE-L:** Measures the longest common subsequence, capturing sentence structure and order.

WHY ROUGE?

- ROUGE is widely accepted for evaluating the quality of summaries by measuring content overlap between the predicted and reference summaries.
- It is particularly effective for extractive summarization, where the goal is to select sentences that are highly representative of the original document.

RESULTS

ROUGE-1

40.3%

- Moderate Overlap In terms of Unigrams
- Suggests that the model is effectively selecting sentences with the important keywords

ROUGE-2

17.5%

- Less Overlap for Bigrams
- The model struggles to capture more complex relationships between words

ROUGE-L

25%

- Moderate Overlap of Sentence Structure
- Captures some sentence level coherance, but may miss longer, more complex sentences

CONCLUSION

- The SummaRuNNer model shows promise for extractive summarization using a sequence-to-label architecture with GRU-based RNNs.
- The model achieves moderate performance with respect to ROUGE scores, especially for uni-grams (ROUGE-1).
- Further refinement can focus on capturing relationships between sentences and improving the model's fluency.

- KANISHK DAD
- DHRUV LIMBAD
- DHYEY PATEL
- AARZOO KHAMBHOO

Q & A