**INTRODUCTION**

Single document summarization is the task of producing a shorter version of a document while preserving its main information content.

Extractive summarization creates a summary by identifying (and subsequently concatenating) the most important sentences in a document.

CNN/DailyMail Story Highlights Dataset (Hermann et al., 2015)

**Contributions**

✓ Optimizing Evaluation Metric: We combine the maximum-likelihood cross-entropy loss with rewards from policy gradient reinforcement learning to directly optimize the evaluation metric, namely ROUGE (Lin and Hovy, 2003).

✓ Learning to Rank: Global optimization framework renders extractive models better at discriminating among sentences for the final summary; a sentence is ranked high for selection if it often occurs in high scoring summaries.

✓ Our model outperforms state-of-the-art extractive and abstractive systems when evaluated automatically and by humans.

**Sentence Ranking with Reinforcement Learning**

Pitfalls of Cross-Entropy Loss

\[ L(\theta) = -\sum_{i=1}^{n} \log p(y_i | s_i, D, \theta) \]

⇒ Misrepresentation of the task with individual (each sentence is assigned 1/0 individually, Cheng and Lapata, 2016) and COLLECTIVE (only few sentences are assigned 1, rests are ignored, Nallapati et al. 2017) labels.

⇒ Maximum-likelihood objective will fail to generalize, as it enforces hard classification.

Reinforcement Learning to Optimize ROUGE

\[ L(\theta) = -E_{y \sim \mu} [r(y)] \]

Policy Gradient with REINFORCE

\[ \nabla L(\theta) = -E_{y \sim \mu} [r(y) \nabla \log p(y|D, \theta)] \]

Training with High Probability Samples

\[ \nabla L(\theta) \approx -r(y) \sum_{i=1}^{n} \nabla \log p(y_i | s_i, D, \theta) \]

**Informativeness Evaluation with Question Answering**

Can participants answer questions (created based on the gold summary) by reading system summaries alone?

Q1: Who backtracked on saying crew reported a pressurization problem? (FAA)
Q2: How many passengers lost consciousness in the incident? (Two)
Q3: How far did the plane descend in three minutes? (28,000 feet)

**Extractive Summarization with Reinforcement Learning**

We use a hierarchical encoder-decoder model to rank sentences in the document for their extract-worthiness. A candidate summary is assembled from the top ranked sentences; the REWARD generator compares the candidate against the gold summary to give a reward which is used in the REINFORCE algorithm (Williams, 1992) to update the model.

**Automatic Evaluation with ROUGE**

**Preference Evaluation**

**Refresh Release**