Rewards with Negative Examples for Reinforced Topic-Focused Abstractive Summarization

Khalil Mrini  
University of California, San Diego  
La Jolla, CA 92093  
khalil@ucsd.edu

Can Liu  
Amazon Alexa  
Seattle, WA 98101  
liuca@amazon.com

Markus Dreyer  
Amazon Alexa  
Seattle, WA 98101  
mddreyer@amazon.com

Abstract
We consider the problem of topic-focused abstractive summarization, where the goal is to generate an abstractive summary focused on a particular topic, a phrase of one or multiple words. We hypothesize that the task of generating topic-focused summaries can be improved by showing the model what it must not focus on. We introduce a deep reinforcement learning approach to topic-focused abstractive summarization, trained on rewards with a novel negative example baseline. We define the input in this problem as the source text preceded by the topic. We adapt the CNN-Daily Mail and New York Times summarization datasets for this task. We then show through experiments on existing rewards that the use of a negative example baseline can outperform the use of a self-critical baseline, in ROUGE, BERTSCORE, and human evaluation metrics.

1 Introduction
Topic-focused summarization is the task of generating a summary given a source text and a specific query or topic. Approaches to topic-focused summarization include query relevance and importance (Gupta et al., 2007), multi-modality manifold ranking (Wan et al., 2007; Wan, 2008; Wan and Xiao, 2009), and query attention (Nema et al., 2017). The DUC 2005 and 2006 datasets (Dang, 2005, 2006) are examples of datasets that are widely used for this task. These datasets are much smaller than benchmark datasets for generic summarization, resulting in fewer research works to train topic-focused summarization on state-of-the-art systems (Deutsch and Roth, 2019).

In parallel, there has been growing work in recent years on reinforcement learning approaches to (generic) abstractive summarization. Proposed rewards aim to optimize non-differentiable summarization metrics like ROUGE (Lin, 2004) and BERTSCORE (Zhang et al., 2019), or to encourage desirable summary aspects like semantic cohesion (Celikyilmaz et al., 2018) and entity coherence (Sharma et al., 2019). Many reinforced abstractive summarization methods use the self-critical baseline or SCST (Rennie et al., 2017) to cap their rewards. This self-critical baseline is obtained by greedily searching for a sequence that maximizes the likelihood probability of the current model.

In this work, we propose a reinforcement learning-based approach to topic-focused summarization. First, we adapt widely used generic summarization benchmarks to this task, such that we aim to generate only one out of three summary sentences, given a corresponding topic. Then, instead of using the self-critical baseline, we introduce a novel baseline that uses negative examples: a sentence that contains information that the summarization model should not focus on.

We run experiments on two existing generic summarization datasets adapted to our task: CNN-Daily Mail (Hermann et al., 2015; Nallapati et al., 2017) and New York Times (Sandhaus, 2008). Our experiments span two existing rewards: the popular ROUGE-L reward and the Distributed Semantic Reward (DSR) of Li et al. (2019), inspired by BERTSCORE (Zhang et al., 2019). Our results show that using our negative example baseline outperforms the self-critical baseline across both datasets and both rewards. We obtain improvements on both datasets in ROUGE and BERTSCORE metrics, and human annotators find that summaries generated with our negative baseline for rewards are generally more relevant to the given topic.

2 Related Work
2.1 Topic-Focused Summarization
There are different definitions of topic-focused summarization. The DUC datasets (Dang, 2005, 2006) propose summarization of documents given a question, also called a query. Vanderwende
et al. (2007) propose SumFocus, a system for topic-focused multi-document extractive summarization. SumFocus is comprised of four components: a generic extractive summarization system, a topic-focused component, sentence simplification, and lexical expansion of topic words.

Deutsch and Roth (2019) define the task of summary cloze as the problem of deciding which content to select in topic-focused summarization, given a context (partial summary). They propose a neural model with separate encoders for the topic and the partial summary.

### 2.2 Reinforcement Learning for Summarization

There is a growing body of work that use reinforcement learning (RL) methods to optimize non-differentiable rewards. ROUGE scores remain a popular RL reward. Other rewards include sentence selection to improve ROUGE scores (Chen and Bansal, 2018; Pasunuru and Bansal, 2018), optimizing question answering metrics (Scialom et al., 2019), and adding desirable custom features to generated summaries (Böhm et al., 2019; Sharma et al., 2019).

Li et al. (2019) find that rewards based on BERTscore (Zhang et al., 2019) and ROUGE each optimize their own metric, but decrease the other one.

Wang et al. (2018) introduce a topic-aware reinforced summarization model. The authors experiment with generic – not topic-focused – summarization datasets, and infer topics using LDA (Blei et al., 2003). Information about the topics is then infused into the model using a topic-aware attention mechanism and topic embeddings.

Whereas a few (Narayan et al., 2018; Li et al., 2018) use the REINFORCE algorithm (Williams, 1992), many RL-based summarization approaches (Paulus et al., 2018; Pasunuru and Bansal, 2018; Li et al., 2018; Celikyilmaz et al., 2018; Yang et al., 2018; Li et al., 2019) use the self-critical sequence training approach (SCST) (Rennie et al., 2017).

### 3 Problem Statement

We tackle the task of topic-focused abstractive summarization as the problem of producing an abstractive summary focused on a given topic, a phrase of one or multiple words. The generated summary should include information from the input text that is related to the topic, and exclude all other information. Consequently, different topics with the same input text should yield different summaries.

More formally, given an input text $x$, a topic $t$ and a corresponding reference summary $y$, we aim to maximize the probability that we generate the right summary:

$$p(y|x, t) > p(y'|x, t)$$

for all $y' \neq y$.

### 4 Rewards with a Negative Example

In self-critical sequence learning (Rennie et al., 2017; Wang et al., 2018), the RL loss formula is as follows for our task:

$$L_{RL} = - (r(y^*) - b) \sum_{i=1}^{N} \log P(y_i^* | y_1^*, \ldots, y_{i-1}^*, x, t)$$

where $b$ is the RL baseline, and $b = r(\tilde{y})$ in self-critical sequence learning. $y^*$ is a sampled summary, and $\tilde{y}$ is a summary obtained greedily by maximizing the probability of the overall sequence.

In our particular task, we propose to use a baseline with a negative example. The intuition is that we encourage the model to generate summaries that are more similar to the reference summary than the negative example. This negative example is an independent sentence from the summary of the corresponding source text, but which does not contain the topic. The negative example acts as a sample of undesirable information, and helps the summarization model learn what kind of information to exclude.

Given a reference summary $y$, and a negative summary $\tilde{y}$, our RL loss with negative examples is defined as in equation 2, where we define the RL baseline: $b = r(\tilde{y})$.

#### 4.1 Rewards

We apply our method on two popular rewards for summarization: rewards based on the ROUGE and BERTSCORE metrics between the sampled summary and the reference summary.

**ROUGE-L reward.** Given a sampled summary $y^*$ and a reference summary $y$, we define the ROUGE-L reward as follows:

$$r_R(y^*) = \text{ROUGE}(y^*, y)$$

**BERTSCORE Reward.** We adopt the Distributed Semantic Reward (DSR) definition of Li et al.
This reward measures the semantic similarity with the reference summary. It is defined as follows:

\[ r_s(y^a) = F_{\text{BERT}}(y^a, y) \]  

(4)

where \( F_{\text{BERT}} \) is the F1 formula of BERTScore (Zhang et al., 2019). It is defined as:

\[ F_{\text{BERT}} = 2 \times \frac{P_{\text{BERT}} \times R_{\text{BERT}}}{P_{\text{BERT}} + R_{\text{BERT}}} \]  

(5)

where the precision \( P_{\text{BERT}} \) and recall \( R_{\text{BERT}} \) are defined as follows for a given reference \( y \) and candidate \( y' \):

\[ P_{\text{BERT}} = \frac{1}{|y'|} \sum_{\tilde{x}_j \in y'} \max_{x_i \in y} x_i^\top \tilde{x}_j \]  

(6)

\[ R_{\text{BERT}} = \frac{1}{|y|} \sum_{x_i \in y} \max_{\tilde{x}_j \in y'} x_i^\top \tilde{x}_j \]  

(7)

4.2 Loss Formula

Whereas Pasunuru and Bansal (2018) train by alternating multiple rewards, Li et al. (2019) propose a single loss formula combining DSR and ROUGE rewards. However, their results show that combining DSR and ROUGE does not yield better results in either ROUGE or \( F_{\text{BERT}} \) scores, compared to using only one reward at a time, along with the summarization loss. We decide to also use only one reward at a time.

We aim to optimize the following loss function:

\[ L = -(1 - \gamma) \times \log p(y|x, t) + \gamma \times L_{RL-N}(y^a, y, \overline{y}) \]  

(8)

where \( \gamma \) is a hyperparameter, and the first term is the negative log-likelihood loss with the reference summary as the target.

5 Experiments and Results

5.1 Datasets

In our experiments, we use two popular summarization benchmarks: the non-anonymized CNN-Daily Mail dataset (Hermann et al., 2015; Nallapati et al., 2017) and the New York Times (NYT) dataset (Sandhaus, 2008).

In the CNN-Daily Mail dataset, summaries are usually 3 independent sentences, corresponding to the “highlights” at the top of the article on the original website. We consider each sentence separately as a reference summary, thereby creating on average 3 datapoints from 1 datapoint in the original dataset.

We filter the large NYT dataset to only get articles with 3-sentence summaries. Similarly to the CNN-Daily Mail dataset, summary sentences are independent, making them fit for topic-focused abstractive summarization.

We use TopicRank (Bougouin et al., 2013), a keyphrase extraction algorithm, to get the 10 most popular keyphrases of the input text. Out of these 10 keyphrases, we pick the highest-scoring one which only appears in the specific summary sentence \( y_i \) to be the topic of \( y_i \). We consider one of the other reference summary sentences of the same input text as the negative example. Therefore, each datapoint in our version of the datasets contains an input text, a reference summary, a negative summary, and a topic. We make sure that no same input text (article) appears in more than one data split. We show the statistics of the dataset used in Table 1.

5.2 Training Details

We set \( \gamma = 0.9984 \) following Paulus et al. (2018) to balance the magnitude difference. We adopt the BART Large architecture (Lewis et al., 2019) as it set a state of the art in the generic summarization of the CNN-Daily Mail dataset, among other tasks. We use a learning rate of \( 3e^{-5} \). We start training from the best model trained on the cross-entropy objective only, on our topic-focused summarization datasets.

We prepend the topic and a separator token to the input text of each datapoint. As the average reference summary is shorter (on average 13 tokens), we set the length of generated summaries at test time between 10 and 20 tokens.
Table 2: Summarization experiment results on the test sets with the ROUGE and BERTSCORE rewards. The model with no reward is a summarizer model trained only on the cross-entropy objective.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>CNN-DAILY MAIL</th>
<th>NEW YORK TIMES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extractive Summarization Baseline Models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1 F1</td>
<td>R2 F1</td>
<td>RL F1</td>
</tr>
<tr>
<td>Lead-1</td>
<td>19.94</td>
<td>6.54</td>
</tr>
<tr>
<td>BM25</td>
<td>22.92</td>
<td>9.71</td>
</tr>
<tr>
<td>SumFocus</td>
<td>32.08</td>
<td>15.82</td>
</tr>
<tr>
<td>Oracle-1</td>
<td>47.52</td>
<td>29.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neural Abstractive Summarization Models</th>
<th>RL b</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>–</td>
<td>38.88 21.91 35.10 76.71 45.84 30.93 40.83 79.32</td>
</tr>
<tr>
<td>ROUGE Self-critical</td>
<td>39.64 22.78 37.10 74.82 46.52 32.43 42.49 77.69</td>
<td></td>
</tr>
<tr>
<td>ROUGE Negative (Ours)</td>
<td>39.97 23.22 37.87 74.12 47.02 32.87 43.12 78.03</td>
<td></td>
</tr>
<tr>
<td>BERTSCORE Self-critical</td>
<td>39.10 21.83 36.77 77.34 45.67 31.62 41.70 79.93</td>
<td></td>
</tr>
<tr>
<td>BERTSCORE Negative (Ours)</td>
<td>39.38 22.41 37.29 77.65 45.51 31.25 41.57 80.35</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Human evaluation ratings of two annotators on 40 sampled summaries from each dataset, comparing reinforced summarization models trained with the negative baseline (ours) vs. the self-critical baseline.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>CNN-DAILY MAIL</th>
<th>NEW YORK TIMES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>REWARD</strong></td>
<td><strong>CRITERIA</strong></td>
<td><strong>Negative</strong></td>
</tr>
<tr>
<td>ROUGE Relevance to Topic</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>Fluency</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>BERTSCORE Relevance to Topic</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>Fluency</td>
<td>16</td>
<td>11</td>
</tr>
</tbody>
</table>

5.3 Baseline Models

We train four extractive baseline models and three abstractive summarization baseline models.

The four extractive baseline models are common summarization baselines, that are meant to give an idea about the difficulty of the task. The first baseline model is Lead-1, which chooses the first sentence from the source article as the generated summary. The second baseline model is BM25 (Robertson et al., 1995), an IR-based score to rank search results given the query. In our case, BM25 ranks sentences of the source article given the topic, and outputs the most relevant sentence as the generated summary. The third baseline model is SumFocus (Vanderwende et al., 2007), an unsupervised probabilistic model for topic-focused summarization. The fourth baseline model is Oracle-1, which greedily searches for the sentence with the highest ROUGE score with the reference summary. This baseline model is meant as an upper-bound for extractive summarization models.

We experiment with three abstractive summarization baseline models that use BART Large. The first baseline model is trained on cross-entropy only. The second and third baseline models are trained with the ROUGE and BERTSCORE (DSR) rewards respectively, with the self-critical method.

5.4 Results and Discussion

We show the results of our experiments in Tables 2. Our proposed approach outperforms the cross-entropy-only BART baseline, but also the two self-critical approaches across both datasets. This shows that negative examples are a good reward baseline in topic-focused summarization. We notice a significant jump in performance in ROUGE-L F1 especially (about 2.5 points), and an increase in ROUGE-1 and ROUGE-2 as well, compared to the cross-entropy-only baseline. Our BERTSCORE-rewarded model achieves the highest FBERT scores, with a slight increase from its self-critical counterpart.

We hire two annotators to judge the fluency and topic relevance of the 40 sampled summary pairs, and therefore get 80 evaluations for each criterion. We ask the two annotators to compare our models with their self-critical counterparts. The annotators are not informed about which model generated
which summary. Results in Table 3 show that our model’s summaries are generally more relevant to the topic, and that our BERTSCORE models are more fluent.

6 Conclusions

We propose a deep reinforcement learning approach to topic-focused abstractive summarization, where we aim to generate summaries focused on a given phrase of one or multiple words. We introduce a new baseline for rewards, based on negative examples collected from independent summary sentences. We show through experiments that our proposed approach outperforms the baseline of self-critical reinforcement learning in the optimized reward metric, and human annotators find our model generates summaries that are more relevant to the topic.

References


Adrien Bougouin, Florian Boudin, and Béatrice Daille. 2013. Topicrank: Graph-based topic ranking for keyphrase extraction.


the Association for Computational Linguistics (Volume 1: Long Papers), pages 1063–1072.


