Generating topic-oriented summaries using neural attention

Kundan Krishna
Adobe Research
kunkrish@adobe.com

Balaji Vasan Srinivasan
Adobe Research
balsrini@adobe.com

Abstract

Summarizing a document requires identifying the important parts of the document with an objective of providing a quick overview to a reader. However, a long article can span several topics and a single summary cannot do justice to all the topics. Further, the interests of readers can vary and the notion of importance can change across them. Existing summarization algorithms generate a single summary and are not capable of generating multiple summaries tuned to the interests of the readers. In this paper, we propose an attention based RNN framework to generate multiple summaries of a single document tuned to different topics of interest. Our method outperforms existing baselines and our results suggest that the attention of generative networks can be successfully biased to look at sentences relevant to a topic and effectively used to generate topic-tuned summaries.

1 Introduction

Automatic text summarization is the task of generating/extracting short text snippet that embodies the content of a larger document or a collection of documents in a concise fashion. Traditionally, researchers have used extractive methods for summarization - where a set of sentences is selected from an article and concatenated *as-is* to form the summary. Extractive methods are limited by their inability to paraphrase the content of the article using new sentences and hence are very different from the summaries created by humans, who paraphrase a given article to generate summaries. Recent works on summarization have come up with sequence-to-sequence models for generating summaries in a word-by-word fashion, thus 'generating' new sentences.

As articles get longer, it might span several topics and therefore, a single summary is often

insufficient to satisfy the interests of the reader. In these cases, it is desirable to produce summaries that are aligned to the topical interests of reader to enable better consumption. Extractive summarization uses sentence-level features (Yang et al., 2017) that have been leveraged for producing query-focused or topic-based summaries. However, for RNN-based frameworks, such tuned summary generation is non-obvious due to the absence of explicit content based features.

To extend the advantages of abstractive summarization and generate topic-tuned summaries, we propose a neural encoder-decoder based framework which takes an article along with a topic of interest as input and generates a summary tuned to the target topic of interest. Our method works by training a neural framework to pay higher attention to parts of the input articles relevant to given topic. To overcome the lack of dataset containing articles with multiple topic-oriented summaries, we use a novel approach to artificially create a topic-centric training corpus from the CNN/Dailymail dataset (Hermann et al., 2015; Nallapati et al., 2016) for training our model.

Table 1 shows an article with different summaries generated by the proposed approach tuned towards the topics of politics, finance and social aspects. It can be seen that the *business* oriented summary talks about IMF's estimates about increase in taxes and the summary for *politics* talks about how it relates to the government's upcoming budget. The summary for the *social* topic elaborates on universal basic income and social security. Thus, for the same article, the proposed approach is capable of generating different summaries tuned to the input topic of interest.
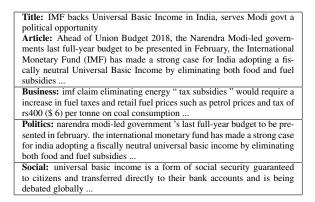
Title: IMF backs Universal Basic Income in India, serves Modi govt a political opportunity
Article: Ahead of Union Budget 2018, the Narendra Modi-led governments last full-year budget to be presented in February, the International Monetary Fund (IMF) has made a strong case for India adopting a fiscally neutral Universal Basic Income by eliminating both food and fuel subsidies ...
Business: imf claim eliminating energy " tax subsidies " would require a increase in fuel taxes and retail fuel prices such as petrol prices and tax of rs400 ($ 6) per tonne on coal consumption ...
Politics: narendra modi-led government 's last full-year budget to be presented in february. the international monetary fund has made a strong case for india adopting a fiscally neutral universal basic income by eliminating both food and fuel subsidies ...
Social: universal basic income is a form of social security guaranteed to citizens and transferred directly to their bank accounts and is being debated globally ...

Table 1: Topic oriented summaries generated by our method for an article (from LiveMint) touching multiple topics

2 Related Work

Traditional methods for summarization (Nenkova and McKeown, 2011) extract key sentences from the source text to construct the summary. Early works on abstractive summarization were focused on sentence compression based approaches (Filippova, 2010; Berg-Kirkpatrick et al., 2011; Banerjee et al., 2015) which connected fragments from multiple sentences to generate novel sentences for the summary or template based approaches that generated summaries by fitting content into a template (Wang and Cardie, 2013; Genest and Lapalme, 2011).

With the advent of deep sequence to sequence models which generated text word-by-word (Sutskever et al., 2014), attention based neural network models have been proposed for summarizing long sentences. Rush et al. (2015) first demonstrated the use of neural networks to generate shorter forms of long sentences. Nallapati et al. (2016) proposed a neural approach for abstractive summarization of large articles by applying the sequence to sequence model. See et al. (2017) further improved the performance on abstractive summarization of articles by introducing the ability to copy words from the source article (Gulcehre et al., 2016) using a pointer network (Vinyals et al., 2015), in addition to generating new words. However, all these frameworks focus on generating a single summary, and do not support topic-tuned summary generation. We use the architecture by See et al. as the starting point for our work and develop a method to generate topic-tuned summaries.

There have been some works on extending extractive summarization towards topical tuning. Lin

and Hovy (2000) proposed the idea of extracting topic-based *signature terms* for summarization. Given a topic and a corpus of documents relevant and not relevant to the topic, a set of words characterizing each topic is extracted using a log-likelihood based measure. Sentences which contain these chosen words are assigned more importance while summarizing. Conroy et al. (2006) further extended the method for query-based multi-document summarization by considering sentence overlap with both query terms and topic signature words.

However, all these works rely on identifying sentence level features to compute topic affinities that are leveraged for choosing topic specific sentences for the summary. Since sequence-to-sequence frameworks generate text in a word-by-word fashion, incorporating sentence level statistics is not feasible in this framework. Therefore, we modify the attention of the network to focus on topic-specific parts of the input text to generate the tuned summaries.

3 Topic aware pointer-generator network

Our method builds on top of the pointer-generator network of See et al. which consists of an encoder and a decoder both based on LSTM architecture. Our contribution is a modified encoder to encode the article in a topic-sensitive manner. However, for the sake of completeness, we shall provide an overview of the entire network, and will reuse notations from their work to a large extent.

Given an input article as a sequence of words $a = w_1, w_2 \ldots w_n$, and a scaled one-hot topic vector u_t, where t is a topic in a predefined set of topics T, the network produces a summary $s = s_1, s_2, \ldots, s_k$ pertaining to the topic t. The topic-vector u_t has a non-zero value $c > 0$ only for the entry corresponding to the desired topic t of the summary. c designates the amount of bias that should be put towards the topic while generating the summary, higher values suggesting higher bias. The input article is passed through an embedding layer to transform them to lower-dimensional embedding $m_1, m_2, \ldots, m_n$. The topic one-hot vector u_t is concatenated to each of the embedding to create the sequence $(m_1, u_t), (m_2, u_t), \ldots, (m_n, u_t)$. This sequence is passed to a bidirectional LSTM encoder which computes a sequence of hidden states $h_1, h_2, ..h_n$. The final hidden state is passed to a decoder which

is also a LSTM, and it serves as the initial hidden state of the decoder. At each decoding step of the decoder, an attention distribution a^t is calculated over all words of the input article,

$$
\begin{aligned}
e_i^t &= v^T \tanh(W_h h_i + W_s s_t + b_{att}) \quad (1) \\
a^t &= \text{softmax}(e^t) \quad (2)
\end{aligned}
$$

Here, s_t is the hidden state of the decoder at timestep t, and v, W_h, W_s and b_{att} are model parameters.

The attention model can be thought of as a probability distribution over words in the source text, which aids the decoder in generating the next word in the summary using words in the source text that have higher attention. Since the encoder has information about what is the topic of interest, the h_i are calculated in such a way that the corresponding areas of the article receiving high attention are relevant to the topic. We demonstrate this later in Section 4.2. The decoder uses this attention to calculate a context vector h_t^* as a weighted sum of the encoder hidden states to determine the next word that is to be generated.

$$
h_t^* = \sum_{i=1}^{n} a_i^t h_i, \quad (3)
$$

At each decoding time step, the decoder network also gets as input y_t, the last word in the summary generated so far and computes a scalar p_{gen} denoting the probability of the network generating a new word from the vocabulary.

$$
p_{gen} = \sigma(w_h^T h_t^* + w_s^T s_t + w_y^T y_t + b_{gen}) \quad (4)
$$

where w_h, w_s, w_y, b_{gen} are trained vectors. The network probabilistically decides based on p_{gen}, whether to generate a new word from the vocabulary or copy a word from the source text using the attention distribution.

For each word w in the vocabulary, the model calculates $P_{vocab}(w)$, the probability of the word getting newly generated next. P_{vocab} is calculated by passing a concatenation s_t and h_t^* through a linear transformation with sigmoid activation. Also, for each word w' in the input article, its total attention received yields its probability of being copied. Since some words occur in the vocabulary and also the input article, they will have non-zero probabilities of being newly generated as well as being copied. Therefore, the total probability of w being

the next word generated in the summary, denoted by **p** is given by,

$$
\mathbf{p}(w) = p_{gen} P_{vocab}(w) + (1 - p_{gen}) \sum_{i:w_i = w} a_i^t \quad (5)
$$

The second term allows the framework to choose a word to copy from the input text using the attention distribution.

The pointer-generator network further employs a coverage mechanism which modifies the loss function to encourage diversity in attention distributions over time steps. This prevents the network from repeating the same phrases again while generating a summary. To handle out-of-vocabulary words, the pointer-generator network also appends them to the vocabulary before the article is encoded.

The training loss is set to be the average negative log-likelihood of the ground truth summaries. The model is trained using back-propagation and the Adagrad gradient descent algorithm (Duchi et al., 2011).

Alternate to the the proposed architecture, one can append the topic vector to the initial hidden state of the encoder, or the initial state of the decoder. However, in our experiments, these approaches did not produce the desired tuning. Another alternative would be to use different W_h attention matrix for each different topic and learn topic specific attention biases. This again did not perform well, perhaps due insufficient data available to train each topic's W_h.

3.1 Generating the training data

The problem of generating *multiple* different summaries of the *same* document based on topics of interest requires a set of articles, each of them (a) annotated with multiple (u_t, s) pairs, where t is a topic, and s is the summary of a oriented to topic t. However, existing datasets for single-document summarization (Over et al., 2007; Hermann et al., 2015) do not have multiple ground truth summaries oriented to different topics for each document. Moreover, existing datasets for topic-based summarization are focused on multi-document summarization and contain multiple articles tagged with a relevant topic. For example, DUC 2005 dataset (Dang, 2005) for topic based multi-document summarization contains 50 topics with each topic having 25 to 50 documents in it. To address the lack of datasets with multiple summaries of a single document, we propose

an approach create such a dataset artificially. The dataset is created in two steps.

We begin with fixing the set of topics for consideration and learn a word frequency based representation for each topic by using a corpus of articles where each article is labelled with a topic. For our experiment, we use the dataset of news articles tagged with topics like *politics*, *sports*, *education* etc. released at the 2017 KDD Data Science + Journalism Workshop (VoxMedia). We group the articles by the topics, so that all articles having topic t are represented by the set S_t. We represent each topic t as a vector $e_t = (n_1, n_2, \ldots, n_v)$ where $v = |V|$ is the size of vocabulary of words $V = \{w_1, w_2, ..., w_v\}$ and n_i is the number of times word w_i occurs in S_t. We normalize e_t to sum to 1.

We then take a corpus of articles with human generated summaries as a collection of (a, s) pairs. For our purposes, we use the CNN-Dailymail dataset which has articles with summaries. We create an *intermediate* dataset which consists of (a, u_t, s) pairs, where u_t is a one-hot representation of the topic t of summary s. We begin with labeling each summary with a topic by computing the dot-product between summary (in its bag-of-words representation) and the topic vectors extracted in the previous step. Let $< v_s, e_t >$ indicate the dot-product between the bag of words representation of summary s and the topic vector for topic t. For each topic t_i, $sim(s, t_i) =< v_s, e_{t_i} >$ is the similarity of summary s to the topic t_i. Let's say that t_i has the highest similarity to summary s and t_j has the second highest similarity. Then we say that summary s has topic t_i with confidence $c = \frac{sim(s, t_i)}{sim(s, t_j)}$. If the confidence is less than a threshold (set to 1.2 in our experiments), we drop the article and summary from our dataset. This enures that the intermediate dataset does not include summaries with more than one dominant topic. If the confidence is greater than the threshold, we add the triple (a, u_{t_i}, s) to the intermediate dataset, where the vector u_{t_i} keeps the value corresponding to t_i equal to the confidence c, instead of 1 as commonly done in one-hot vectors. This allows us to retain the confidence of s being of topic t_i while training our model.

To generate the final dataset, we follow the steps below:

1. Randomly pick (a_1, u_{t_1}, s_1) and (a_2, u_{t_2}, s_2) from the intermediate dataset such that $t_1 \neq t_2$.

2. Make a new article a' by sequentially picking up lines from a_1 and a_2. Each addition of a new line is done by randomly selecting one of a_1 or a_2 and extracting out a new line from the beginning of it. This ensures that the lines from a_1 occur in the same order in a' as originally in a_1, and the same thing is true for a_2 too. This ensures that the sequential flow of content is retained in the merger.

3. Add (a', u_{t_1}, s_1) to the final dataset.

4. Repeat step 2 to get a new article a'' and add (a'', u_{t_2}, s_2) to the final dataset.

5. Discard (a_1, u_{t_1}, s_1) and (a_2, u_{t_2}, s_2) from the intermediate dataset.

6. Repeat steps $1-5$ until the entire intermediate dataset is exhausted or all remaining instances in it have the same topic.

The created final dataset is used to train the proposed neural network for summarization. Since every article in the final dataset is a combination of two original articles, and the target summary to be generated is of one of them, the model must learn to distinguish between content coming from the two original articles. The randomization of position of sentences from each original article while merging, ensures that there is no position-specific bias that the model can use either. Since the two original articles have different topics, and the only information given to the model to hint whose summary is to be generated is the topic of one of them, the model is forced to learn what content is more relevant to the given topic and generate a summary accordingly.

We ended up with $112, 360$ articles in our final dataset since many article-summary pairs from the CNNDailymail dataset were dropped due to insufficient confidence about their topics. Out of this, $103, 666$ were used for training, $4, 720$ for validation, and $3, 974$ for testing.

4 Experimental Evaluation

To position our method against existing works, we use the following summarizers as our baselines.

Our first baseline is the **vanilla pointer generator (PG)** described in the original work of See et al. (2017). This method does not consider the

desired topic of summary when generating a summary. For an unbiased evaluation, we use exactly those unmerged article-summary pairs of the CNN/Dailymail dataset for training and validation which were eventually incorporated in the final dataset. Then the trained model is applied to generate summaries of the test set of the final dataset.

Our next baseline is a **frequency-based extraction** method that selects lines from the input article which are strongly aligned to the desired topic u_t. For each sentence, the relevance to each of the predefined topics is calculated using a dot product between their vector representations. The sentence is designated to the topic having the maximum relevance score, and the strength of alignment is the ratio of the the highest and the second highest relevance scores.

We extract all the sentences in the article which are aligned to the target topic, and run it through the pointer-generator network to create the summary. We refer to this baseline as **abstractive summarizer with frequency based extraction (*Freq-Abs*)**. Alternatively, we take k sentences which have the highest strength of alignment with the target topic to create a purely extractive summary. k was set to 3 in accordance with the average number of sentences in the summaries of the training set (2.83). We call this method **extractive summarizer with frequency based extraction (*Freq-Ext*)**.

Our last baseline is a **topic-signature based approach** which also works by extracting sentences from the article which are aligned to the target topic. However, the selection of sentences is based on topic signatures as described by Lin and Hovy (2000) and Conroy et al. (2006) instead of word frequencies. A topic signature is a set of words relevant to the topic. For any given sentence, the number of signature terms of each topic is computed. The sentence is designated to belong to the topic which has the highest number of its signature terms occurring in it.

The topic signature is determined based a set of documents T relevant to the topic, and a set of background documents T' that is indicative of general topics. It is assumed that in T and T', the occurrence of each word w follows a binomial distribution with probability of occurrence p. The likelihood of observing T and T' is calculated under two hypotheses - one where the probability of occurrence of w is p_1 in T and p_2 in T' such that

$p_1 > p_2$, and the other where it is p in both T and T'. The ratio of likelihoods is calculated and words for which this ratio is the highest are included as part of the topic's signature.

We extracted topic signatures using the summaries of the training dataset as our corpus. For each topic t, the corresponding summaries form the topic specific corpus T, and the remaining summaries make the background corpus. Table 2 shows a subset of the topic signatures.

Topic	Signature words
business	stock, firm, rate, google, worth, companies
education	children, teacher, parents, schools, test
entertainment	actor, star, league, movie, will, taylor
health	disease, brain, people, risk, league, infection, blood
military	arrested, police, car, officer, year, officers, killed, charged
politics	party, campaign, secretary, democrats, bill, congress
social	church, rich, life, woman, sex, identity, gender, will
sports	scored, united, league, nfl, cup, england, beat, game
technology	app, solar, launch, ipod, app, earth, energy, oil

Table 2: Words from signatures of different topics

Analogous to Freq-Abs and Freq-Ext, we try two alternatives here as well - **abstractive summarizer with signature based extraction (*Sign-Abs*)** and **extractive summarizer with signature based extraction (*Sign-Ext*)**.

4.1 Performance on the created dataset

We used the $3,974$ article-topic-summary tuples from our final dataset to evaluate the performance of the summarizers. The models were given the input article and topic and the generated summary was compared with the ground truth summary. We use ROUGE scores to measure the quality of summaries. ROUGE scores measure the precision, recall and f-measure for the occurrence of n-grams in the generated summary with respect to the reference human generated summary. We use the ROUGE-1, ROUGE-2 and ROUGE-L variants (Lin and Och, 2004) which look at unigram, bigram and longest common subsequence overlaps between generated and reference summaries. Table 3 shows the ROUGE F1 scores for the topic-based summaries generated by different methods. It is easy to see that the proposed method yields the best performance across all the baselines.

Further, we also observed that the summaries generated by our system show abstractive nature as noted in See et al. (2017). Table 4 shows some instances where our model used new words unseen in the article.

Algorithm	ROUGE-1	ROUGE-2	ROUGE-L
PG	26.8	9.2	24.5
Freq-Abs	25.8	8.4	23.4
Freq-Ext	25.5	8.5	22.9
Sign-Abs	26.1	8.5	23.7
Sign-Ext	25.9	8.7	23.3
Our method	**34.1**	**13.6**	**31.2**

Table 3: ROUGE F1 scores obtained by various methods on the final test set

Article: spain 's 2-0 defeat by holland on tuesday brought back bitter memories of their disastrous 2014 world cup , but coach vicente del bosque will not be too worried ...   **Summary:** holland **beat** spain 2-0 at the amsterdam arena on tuesday **night**
Article: it 's 11 years since arsenal won the title. they went from invincibles to incapables ...   **Summary:** arsene wenger 's **side** have 15 wins in 17 **appearances**

Table 4: Summaries where our model uses new words not seen in the input article

4.2 Performance on multi-topic articles

We ran the proposed model on original articles from CNN-DailyMail dataset which were not part of the final dataset used for training. Since the articles were not annotated with topics, we generated summaries for all the different topics. We then detected articles where the method generated different summaries for different topics suggesting the presence of more than one topic in the article. Table 5 shows a few summaries generated by the proposed approach where different summaries were generated aligned to the input topics. The first article talks about the dropping of a player from a football squad. The summary oriented to the *military* topic talks about the assault of a police officer by the player and his criminal history. The *sports* summary talks about the player's fate in the remaining games of the season.

Similarly, we have different summaries for the second article, where the *education* oriented summary talks about the educational affiliations of the suspects and disciplinary procedures, whereas the *military* summary talks about the arrests. Note that these are among the original articles in the dataset which were not used for creating any of the articles in our final data used for training the model.

We also observe that the attention distribution

Title: Paul McGowan won't be risked in final six games of the season, as Dundee boss Paul Hartley looks to help troubled midfielder   **Military:** dundee rogue paul mcgowan has been handed his third conviction after assaulting a police officer . mcgowan escaped a jail sentence but was placed under a restriction ...   **Sports:** paul hartley has warned that paul mcgowan may not feature in any of the team 's remaining six games this season because he will not risk playing the troubled midfielder this season ...
Title: Third suspect arrested in alleged Panama City gang rape   **Education:** ryan calhoun has been a student at middle tennessee state university . the two are students and have been " placed on temporary suspension and disciplinary procedures ...   **Military:** sheriff 's office : third person has been arrested in the case of an alleged spring break gang rape that was videotaped on a crowded stretch of panama city beach , the bay county , florida , sheriff 's office said . the arrests come after a woman told police ...
Title: University of Michigan will go forward with 'American Sniper' screening   **Military:** kyle was fatally shot at a texas shooting range in 2013 . kyle cooper was nominated for an oscar for his portrayal of chris kyle , a navy seal and most lethal sniper in u.s. military history ...   **Education:** university of michigan has decided to proceed with a screening of the film " american sniper " despite objections from some students . more than 200 signed a petition ...
Title: The Pope's old iPad sells for $30,500   **Business:** the motorcycle sold for $ 284,000 at auction , more than 10 times its normal sales price . a harley motorcycle jacket signed by francis sold for nearly $ 68,000 . ...   **Education:** ... the proceeds will go to a school in montevideo , uruguay . it 's not the first time a papal hand-me-down has gone for big bucks . ...

Table 5: Topic oriented summaries generated by our method for articles from CNN-Dailymail dataset

varies according to the input topic of summary. We investigate the amount of cumulative attention (defined as coverage by See et al.) that is received by each term summed over all the decoding steps during generation. An example of the variation in attention over an input article is shown in Figure 1. The degree of yellow hue is used to represent the value of coverage for each term. For the topic *military*, words like *jail* and *assaulting* receive higher attention, whereas for the topic *sports* words like *games* and *player* are highlighted. Figure 1(b) also shows an example where our model summarizes by skipping a clause (*"which kick off at 3pm"*).

4.3 Human evaluation of performance

Finally, we performed human evaluations of summaries to compare the quality of our method against the baselines. We restrict ourselves to

dundee boss paul hartley has warned that paul mcgowan may not feature in any of the team 's remaining six games this season because he fears playing the troubled midfielder could see him break his police curfew and end up in prison . after his third conviction last week for assaulting a police officer , mcgowan escaped a jail sentence but was placed under a restriction of liberty order and confined to his home from 7pm to 7am for 16 weeks . that has already ruled him out of the evening matches with celtic at dens park on wednesday and at parkhead on friday may 1 , but hartley fears that even playing in the three remaining games in the city of discovery , which kick off at 3pm , could leave the player at risk of breaching his order . dundee rogue paul mcgowan has been handed his third conviction after assaulting a police officer . dundee manager paul hartley

(a) Coverage for the topic *military*

dundee boss paul hartley has warned that paul mcgowan may not feature in any of the team 's remaining six games this season because he fears playing the troubled midfielder could see him break his police curfew and end up in prison . after his third conviction last week for assaulting a police officer , mcgowan escaped a jail sentence but was placed under a restriction of liberty order and confined to his home from 7pm to 7am for 16 weeks . that has already ruled him out of the evening matches with celtic at dens park on wednesday and at parkhead on friday may 1 , but hartley fears that even playing in the three remaining games in the city of discovery , which kick off at 3pm , could leave the player at risk of breaching his order . dundee rogue paul mcgowan has been handed his third conviction after assaulting a police officer . dundee manager paul hartley

(b) Coverage for the topic *sports*

Figure 1: Variation in the attention coverage while summarizing an article for different topics

the best performing baseline - the topic signature based abstractive summarizer (Sign-Abs), and the vanilla pointer-generator.

We fetched articles touching multiple topics using the NYTimes Search API[1], which allows to search for articles on NYTimes which appear in the news desk for a topic (the major topic) and are tagged with another topic (the minor topic). For different pairs of topics, we retrieved relevant articles using the API and randomly selected few of them to generate summaries tuned to the two topics using our method and the baselines. We retrieved 18 total articles for the evaluation. Each article had two topics leading to 36 (article,topic) pairs for the summary generation. For each (article,topic) pair, annotators were shown two summaries - one generated by our method and the other by one of the baselines. The task was to choose the summary more relevant to the topic. Each such task was annotated by 10 different annotators. Every annotator was assigned 9 tasks and 1 extra dummy task to check if they were paying attention. Annotations from evaluators who answered incorrectly to the dummy task were discarded. We had a total of 720 annotations from the human evaluation and their summary is shown in Table 6.

The value under "Overall annotations" refers to the fraction of all human responses across all

documents which rated the summary produced by our method better than the alternate summary produced by the compared baseline method. Understandably, the proposed approach is comparable in the preference to both the baselines for the major topic (0.5667) - since most standard summaries will cover the primary topic of the input. However, for the minor topic, it can be seen that the proposed approach is better than the baselines.

The value under "Document annotations" indicates the fraction of times the proposed method was preferred by half or more of the annotators for a document-topic pair. It is easy to see that the proposed approach clearly outperforms the baselines under this scenario. The difference in performance is even more significant for summaries generated for the non-major topic since our approach is capable of efficiently generating tuned summaries for minor topics as well.

5 Conclusion

We proposed a method for generating multiple abstractive summaries of a given document oriented towards different topics of interest. Our method works by modifying the attention mechanism of a pointer-generator neural network to make it focus on text relevant to a topic. For training our network, we devised a novel way to create a dataset where articles are tagged with topic oriented summaries. Our method outperformed previous fea-

Topics	Overall Annotations		Document Annotations	
	vs PG	vs Sign-Abs	vs PG	vs Sign-Abs
All	0.5889	0.6111	0.7222	0.8333
Major	0.5667	0.5667	0.6667	0.7778
Minor	0.6111	0.6556	0.7778	0.8889

Table 6: Evaluation of summaries of the proposed approach against Pointer Generator Framework and Topic Signature based Summarizer by human annotators

ture based methods for topic oriented summarization using word frequencies or log likelihood ratios.

References

Siddhartha Banerjee, Prasenjit Mitra, and Kazunari Sugiyama. 2015. Multi-document abstractive summarization using ilp based multi-sentence compression. In *IJCAI*. pages 1208–1214.

Taylor Berg-Kirkpatrick, Dan Gillick, and Dan Klein. 2011. Jointly learning to extract and compress. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, pages 481–490.

John M Conroy, Judith D Schlesinger, and Dianne P O'Leary. 2006. Topic-focused multi-document summarization using an approximate oracle score. In *Proceedings of the COLING/ACL on Main conference poster sessions*. Association for Computational Linguistics, pages 152–159.

Hoa Trang Dang. 2005. Overview of duc 2005. In *Proceedings of the document understanding conference*. volume 2005, pages 1–12.

John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research* 12(Jul):2121–2159.

Katja Filippova. 2010. Multi-sentence compression: Finding shortest paths in word graphs. In *Proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics, pages 322–330.

Pierre-Etienne Genest and Guy Lapalme. 2011. Framework for abstractive summarization using text-to-text generation. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*. Association for Computational Linguistics, pages 64–73.

Caglar Gulcehre, Sungjin Ahn, Ramesh Nallapati, Bowen Zhou, and Yoshua Bengio. 2016. Pointing the unknown words. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. volume 1, pages 140–149.

Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In *Advances in Neural Information Processing Systems*. pages 1693–1701.

Chin-Yew Lin and Eduard Hovy. 2000. The automated acquisition of topic signatures for text summarization. In *Proceedings of the 18th conference on Computational linguistics-Volume 1*. Association for Computational Linguistics, pages 495–501.

Chin-Yew Lin and Franz Josef Och. 2004. Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, page 605.

LiveMint. 2017. IMF backs Universal Basic Income in India, serves Modi govt a political opportunity. [Online; accessed 8-Dec-2017].

Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Ça glar Gulçehre, and Bing Xiang. 2016. Abstractive text summarization using sequence-to-sequence rnns and beyond. *CoNLL 2016* page 280.

Ani Nenkova and Kathleen McKeown. 2011. Automatic summarization. *Foundations and Trends® in Information Retrieval* 5(2–3):103–233.

Paul Over, Hoa Dang, and Donna Harman. 2007. Duc in context. *Information Processing & Management* 43(6):1506–1520.

Alexander M Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. pages 379–389.

Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. volume 1, pages 1073–1083.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. pages 3104–3112.

Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In *Advances in Neural Information Processing Systems*. pages 2692–2700.

VoxMedia. 2017. Data science + journalism @ kdd 2017. https://sites.google.com/view/dsandj2017/datasets. [Online; accessed 8-Dec-2017].

Lu Wang and Claire Cardie. 2013. Domain-independent abstract generation for focused meeting summarization. In *ACL (1)*. pages 1395–1405.

Yinfei Yang, Forrest Sheng Bao, and Ani Nenkova. 2017. Detecting (un) important content for single-document news summarization. *EACL 2017* page 707.