Summarizing Text on Any Aspects: A Knowledge-Informed Weakly-Supervised Approach

Abstract

Given a document and a target aspect (e.g., a topic of interest), aspect-based abstractive summarization attempts to generate a summary with respect to the aspect. Previous studies have assumed a small pre-defined set of aspects and fall short of summarizing on other diverse topics. In this work, we study summarizing on arbitrary aspects relevant to the document, which significantly expands the application of the task in practice. Due to the lack of supervision data, we develop a new weak supervision construction method and an aspect modeling scheme, both of which integrate rich external knowledge sources such as ConceptNet and Wikipedia. Experiments show our approach achieves performance boosts on summarizing real and synthetic documents given pre-defined or arbitrary aspects.

1 Introduction

Remarkable progresses have been made in generating *generic* summaries of documents (Nallapati et al., 2016; See et al., 2017; Narayan et al., 2018), partially due to the large amount of supervision data available. In practice, a document, such as a news article or a medical report, can span multiple topics or aspects. To meet more specific information need in applications such as personalized intelligent assistants, it is often useful to summarize a document with regard to a given aspect, i.e., *aspect-based* summarization.

Recent research has explored the problem of aspect-based abstractive summarization (Krishna and Srinivasan, 2018; Fremann and Klementiev, 2019). A key challenge of the task is the lack of direct supervision data containing documents paired with multiple aspect-based summaries. Previous studies have created synthetic data from generic news summarization corpora which have a small set of aspects (e.g., "sports", "health" and other 4 aspects in (Frermann and Klementiev, 2019)). As a result, models trained on the data tend to be restricted to the pre-defined set and fall short of summarizing on other diverse aspects.

This paper aims to go beyond pre-defined aspects and enable summarization on arbitrary aspects relevant to the document. The arbitrary aspect may not be explicitly mentioned but only implicitly related to portions of the document, and it can be a new aspect not seen during training. To this end, we develop a new approach that integrates rich external knowledge in both aspect modeling and weak supervision construction. Specifically, we derive weak supervisions from a generic summarization corpus, where the ConceptNet knowledge graph (Speer et al., 2017) is used to substantially expand the aspect scope and enrich the supervisions. To assist summarization model to better understand an aspect, especially a previously unseen one, we augment the model inputs with rich aspect-related information extracted from Wikipedia.

Our approach is compatible with any neural encoder-decoder architectures. In this work, we use the large pre-trained BART model (Lewis et al., 2019) and fine-tune with the proposed method. Experiments on real news articles show our approach achieves performance boosts over existing methods. When adapting to the previous synthetic domain (Frermann and Klementiev, 2019), the BART model after fine-tuning with our weak supervisions becomes much more data efficient, and significantly outperforms previous best-performing systems using only 0.4% training examples.

2 Related Work

Early work has studied topic-focused summarization in the multi-document setting, with (typically small) datasets containing multiple documents tagged with a relevant topic (Dang, 2005; Conroy

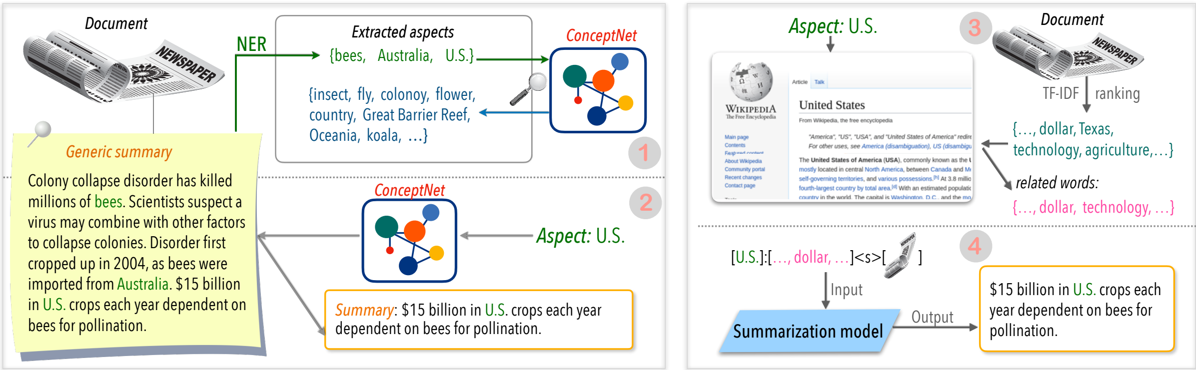

Figure 1: Illustration of our approach. **Left:** Constructing weak supervisions using ConceptNet, including (1) extracting aspects and (2) synthesizing aspect-based summaries. **Right:** Augmenting aspect information, including (3) identifying aspect related words in the document using Wikipedia and (4) feeding both aspect and related words into summarization model.

et al., 2006). For single-document aspect-based summarization, *extractive* methods were used to extract related key sentences/words from the document (Lin and Hovy, 2000). This work studies *abstractive* aspect-based summarization that generates summaries. Deutsch and Roth (2019) studied a sub-task of learning to select information that should be included in the summary. Recent work (Frermann and Klementiev, 2019; Krishna and Srinivasan, 2018) on the problem synthesized training data which use news categories as the aspects and thus have a small pre-defined set of aspects available. We aim to enable summarization on any aspects, and develop new weak supervisions by integrating rich external knowledge.

Aspect-based summarization has also been explored in the customer reviews domain (Hu and Liu, 2004), where product aspects, customer sentiment, and sometimes textual summaries are extracted (Popescu and Etzioni, 2007; Wang and Ling, 2016; Angelidis and Lapata, 2018). Query-based summarization produces a summary in response to a natural language query/ question (Daumé III and Marcu, 2006; Liu et al., 2012; Xie et al., 2020) which differs from abstract aspects.

3 Approach

Given a document and an aspect which can be a word or a phrase, the task aims to generate a summary that concisely describes information in the document that is relevant to the aspect. We present our approach that enables a neural summarization model to summarize on any aspects. The aspect can be any words relevant to (but not necessarily occurring in) the document. Our approach incorporates rich external knowledge sources, including ConceptNet for enriching weak supervisions in training (sec 3.1) and Wikipedia for advising the document-aspect relation to improve comprehension (sec 3.2). Figure 1 shows an overview of our approach.

An advantage of our approach is that it is compatible with any neural summarization architectures, such as the popular encoder-decoders. This enables us to make use of the large pre-trained network BART (Lewis et al., 2019), on which we apply our approach for fine-tuning and improved inference.

3.1 Knowledge-enriched Weak Supervisions

Usually no direct supervision data is available. We start with a generic summarization corpus. Specifically, in this work we use the CNN/DailyMail (Hermann et al., 2015) which consists of a set of *(document, summary)* pairs. Our approach constructs weakly supervised examples by automatically extracting potential aspects and synthesizing aspect-based summaries from the generic summary. Each resulting aspect and its aspect-based summary are then paired with the document for training.

Extracting Aspects Given a generic summary, we want to extract as many aspects as possible so that the summarization model can see sufficient examples during training. On the other hand, the aspects must be relevant to the generic summary to facilitate synthesizing appropriate summary in the next step. To this end, we first apply a named entity recognition (NER) model[1] to extract a set of entities mentioned in the generic summary. These entities serve as a seed set of aspects. We then augment the seed set by collecting each entity's neighbor concepts on the ConceptNet knowledge graph, as these concepts are semantically closely related to the entity (and thus the generic summary). For example, in Figure 1(1), "insect" is a new aspect from ConceptNet given the seed entity "bees".

[1]https://spacy.io/models/xx

Synthesizing Aspect-based Summaries For each aspect, we synthesize a specific summary by extracting and concatenating all relevant sentences from the generic summary. We make use of ConceptNet in a similar way as above. Specifically, a sentence is considered relevant if it mentions the aspect or any of its neighbors on ConceptNet.

The use of ConceptNet greatly augments the supervisions in terms of both the richness of aspects and the informativeness of respective summaries.

3.2 Knowledge-aided Aspect Comprehension

The summarization model is required to precisely locate information in the document that matches the desired aspect. Such comprehension and matching can be challenging, especially with only noisy weak supervisions during training. Our approach facilitates the inference by informing the model with pre-computed document-aspect relations.

Concretely, we extract words from the document which are most *related* to the aspect (more details below), and feed those words into the model together with the aspect and document. In this way, the model is advised which parts of the document are likely to be aspect-related. For the BART architecture, we use an input format as:

`[aspect]:[related words]<s>[doc]`

where `<s>` is a special token for separation.

To determine the related words, the intuition is that the words should be describing or be associated with the aspect. We use the Wikipedia page of the aspect for filtering the words. Besides, we want to select only salient words in the document for a concise summary. Thus, we first rank all words in the document by TF-IDF scores, and select top words that occur in the aspect's Wikipedia page[2].

4 Experiments

All code is attached in the supplementary materials and will be released.

Setup As above, we construct weak supervisions from the CNN/DailyMail dataset (Hermann et al., 2015) which consists of 280K (doc, summary) pairs in training set. We use the CNN/DailyMail-pretrained BART (Lewis et al., 2019) provided by Fairseq (Ott et al., 2019) as our base summarization model, and fine-tune with our approach. We use Adam optimizer with an initial learning rate of 3e-5, and beam search decoding with a width of 4.

[2]We select $\leq$ 10 words. If the Wikipedia API does not find any page of the aspect, the related word is set to empty.

Models	R-1	R-2	R-L
Lead-3 (2019)	21.50	6.90	14.10
PG-Net (2017)	17.57	4.72	15.94
SF (2019)	28.02	10.46	25.36
BART Supervised 280K	41.90	20.46	39.06
BART Supervised 1K	24.58	8.82	22.74
BART Weak-Sup (Ours)	28.56	10.53	25.93
+ Supervised 1K (Ours)	35.62	15.80	33.01

Table 1: Results (ROUGE) on the MA-News test set. The results of `Lead-3`, `PG-Net` and `SF` are from (Fremann and Klementiev, 2019), where `SF` is the previous best model. Our approach trains with only weak supervisions (sec 3.1) or with additional 1K MA-News supervised training data.

4.1 Studies on Synthetic Domain

We first study on the synthetic data, **MA-News**, introduced in (Frermann and Klementiev, 2019). Although its aspects are restricted to only 6 coarse-grained topics, the synthetic domain facilitates automatic evaluation, providing a testbed for (1) comparison with the previous models and (2) studying the generalization ability of our weak-supervision approach when adapting to the new domain.

Specifically, MA-News is synthesized from CNN/DailyMail by interleaving paragraphs of original documents belonging to different aspects. The assembled document is paired with each component's aspect and generic summary to form an aspect-based summary instance. The dataset has 280K/10K/10K examples in train/dev/test sets, respectively, and contains 6 pre-defined aspects including {“sport", “health", “travel", “news", “science technology", “tv showbiz"}.

Comparisons with previous methods We first compare our approach with the previous summarization models, as shown in Table 1. **(1)** In the first block, `SF` is the best model in (Fremann and Klementiev, 2019) with a customized neural architecture and is trained with the full MA-News training set. **(2)** In the second block, we also train the large BART model with the MA-News training set, either using the full 280K instances or only 1K instances. BART trained with the full set unsurprisingly shows much better results than `SF`, yet the one with the 1K subset falls behind `SF`. **(3)** The third block evaluates our method. `BART Weak-Sup` is fine-tuned only with our weak supervisions (sec 3). Even without using any direct supervision examples in MA-News, the model performs slightly better than `SF`. More interestingly, by further using only 1K MA-News instances to continue fine-tuning the model, we achieve performance boosts compared to both `SF`

Models	R-1	R-2	R-L
Weak-Sup only	28.56	10.53	25.93
Supervised 1K	24.58	8.82	22.74
+ Weak-Sup	**35.62**	**15.80**	**33.01**
Supervised 3K	29.13	11.89	27.02
+ Weak-Sup	**37.17**	**16.84**	**34.40**
Supervised 10K	39.49	18.71	36.67
+ Weak-Sup	**39.82**	**18.81**	**36.92**

Table 2: Fine-tuning BART on the synthetic domain, evaluated on MA-News test set. `Weak-Sup only` trains BART only with our weak supervisions. `Supervised 1K` trains with 1K MA-News examples. `+Weak-Sup` trains first with weak supervisions and then supervisedly on MA-News.

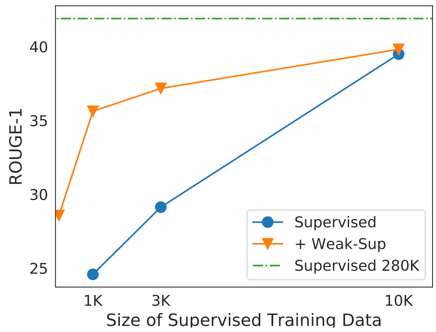

Figure 2: Visualizing the ROUGE-1 results in Table 2. The green dashed line marks the performance of BART fine-tuned on the whole MA-News training set.

and `BART supervised 1K`. This shows our proposed knowledge-informed method provides rich information that helps with the task.

Efficiency of adapting to the domain We continue to study how our weakly supervised method can help with efficient adaptation of BART to the synthetic domain. As shown in Table 2, by fine-tuning BART using more MA-News training data (i.e., `Supervised 1K`, `3K`, and `10K`), the test performance improves reasonably, as is also shown by the blue curve in Figure 2. However, if we add our proposed weak supervisions (i.e., `+Weak-Sup`), the performance improves much faster, as is also shown by the orange curve in the figure. The enhanced data efficiency validates the effectiveness of the weakly supervised method.

4.2 Summarizing on Any Aspects

We next study summarization of a document on arbitrary aspects. To evaluate the generalization of the methods, we test on real news articles from the All The News corpus (Kaggle, 2020) where we randomly extract 50 articles from different publications other than CNN. We ask human annotators to label an arbitrary relevant aspect for each article. We then collect aspect-based summaries by the models, and present each to 3 annotators to rate. As in previous work (Kryscinski et al., 2019; Fremann and Klementiev, 2019), the criteria include accuracy (i.e., summary-aspect relatedness), informativeness, and fluency. The Cohen's Kappa score is 0.51, showing moderate inter-rater agreement. Table 3 shows the results, where we can see our weakly supervised method performs best. The model trained on the 280K MA-News examples, though performs well on the MA-News test set (Table 1), fails to generalize to the broader set of diverse aspects, showing the importance of introducing rich knowledge in supervisions and inference process for generalization. Interestingly, fine-tuning our model with 3K MA-News instances results in inferior performance, showing the previous synthetic data with limited aspects could restrict generalization to other aspects.

Models	Accu.	Informative.	Fluency
MA-News 280K	2.19	3.44	4.44
Weak-Sup (Ours)	**4.59**	**4.36**	**4.87**
+ MA-News 3K (Ours)	4.14	4.07	4.80

Table 3: Human evaluation using 5-point Likert scale. `MA-News 280K` trains BART with the whole MA-News set. `Weak-Sup` trains with our weak supervisions. `+MA-News 3K` further fine-tunes with 3K MA-News instances.

Document: [...] The populist billionaire denounced Clinton's suggested proposal as "crazy" and "unfair" to American workers [...] According to Pew polling data, Hillary Clinton's plan to expand immigration is opposed by at least 83 percent of the American electorate — voters whom Clinton has suggested are racist for opposing immigration. According to a September 2015 Rasmussen survey, 85 percent black voters oppose Clinton's refugee agenda to admit more than 100, 000 Middle Eastern refugees — with less than one percent of black voters (.56 percent) in favor of her refugee plan.

Aspect: vote
Summary: Polls show that at least 83 percent of the U.S. electorate is opposed to expanding immigration and that 85 percent of black voters oppose the plan to admit more than 100,000 middle eastern refugees to the country.

Table 4: Example summary. The aspect "vote" does not occur in the document. Related words (sec 3.2) are in orange.

Table 4 shows a summary by our `Weak-Sup` model, where we can see the identified related words are closely related to the aspect, and the resulting summary is both relevant and informative. We show more examples in the appendix.

5 Conclusions

We have developed a new knowledge-informed weakly supervised method to enable summarizing a document on arbitrary relevant aspects. The promising empirical results motivate to explore more on integrating external knowledge in the task.

References

Stefanos Angelidis and Mirella Lapata. 2018. Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised. In *EMNLP*.

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*, pages 152–159.

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

Hal Daumé III and Daniel Marcu. 2006. Bayesian query-focused summarization. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 305–312. Association for Computational Linguistics.

Daniel Deutsch and Dan Roth. 2019. Summary cloze: A new task for content selection in topic-focused summarization. In *EMNLP*, pages 3711–3720.

Lea Frermann and Alexandre Klementiev. 2019. Inducing document structure for aspect-based summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6263–6273, Florence, Italy. Association for Computational Linguistics.

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.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *KDD*, pages 168–177.

Kaggle. 2020. All the news 2.0: 2.7 million news articles. `https://components.one/datasets/all-the-news-2-news-articles-dataset/`.

Kundan Krishna and Balaji Vasan Srinivasan. 2018. Generating topic-oriented summaries using neural attention. In *NAACL*, pages 1697–1705.

Wojciech Kryscinski, Nitish Shirish Keskar, Bryan McCann, Caiming Xiong, and Richard Socher. 2019. Neural text summarization: A critical evaluation. In *EMNLP*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

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*, pages 495–501.

Yan Liu, Sheng-hua Zhong, and Wenjie Li. 2012. Query-oriented multi-document summarization via unsupervised deep learning. In *AAAI*.

Ramesh Nallapati, Bowen Zhou, Caglar Gulcehre, Bing Xiang, et al. 2016. Abstractive text summarization using sequence-to-sequence rnns and beyond. In *CoNLL*.

Shashi Narayan, Shay B Cohen, and Mirella Lapata. 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *EMNLP*.

Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. fairseq: A fast, extensible toolkit for sequence modeling. In *NAACL*.

Ana-Maria Popescu and Orena Etzioni. 2007. Extracting product features and opinions from reviews. In *Natural language processing and text mining*, pages 9–28. Springer.

Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *AAAI*.

Lu Wang and Wang Ling. 2016. Neural network-based abstract generation for opinions and arguments. In *NAACL*.

Yujia Xie, Tianyi Zhou, Yi Mao, and Weizhu Chen. 2020. Conditional self-attention for query-based summarization. *arXiv preprint arXiv:2002.07338*.