Integrating Semantic Scenario and Word Relations for Abstractive Sentence Summarization

Yong Guan[1], **Shaoru Guo**[1], **Ru Li** [1,2*], **Xiaoli Li**[3], **Hu Zhang**[1,2],
1. School of Computer and Information Technology, Shanxi University, Taiyuan, China
2. Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, China
3. Institute for Infocomm Research, A*Star, Singapore
{guanyong0130,guoshaoru0928}@163.com
{liru,zhanghu}@sxu.edu.cn,xlli@i2r.a-star.edu.sg

Abstract

Recently graph-based methods have been adopted for Abstractive Text Summarization. However, existing graph-based methods only consider either word relations or structure information, which neglect the correlation between them. To simultaneously capture the word relations and structure information from sentences, we propose a novel Dual Graph network for Abstractive Sentence Summarization (DG-ABS). Specifically, we first construct *semantic scenario graph* and *semantic word relation graph* based on FrameNet, and subsequently learn their representations and design graph fusion method to enhance their correlation and obtain better semantic representation for summary generation. Experimental results show our model outperforms existing state-of-the-art methods on two popular benchmark datasets, i.e., Gigaword and DUC 2004.

1 Introduction

Abstractive text summarization is a challenging Natural Language Generation (NLG) task, aiming to compress or rewrite a text into a short version while preserving its essential information. Here, we focus on abstractive sentence summarization where the input text is a sentence (Rush et al., 2015). Traditional methods for text summarization are mainly about the feature-based machine learning methods, such as template methods (Zhou and Hovy, 2004) and syntactic tree pruning (Knight and Marcu, 2002). They are, however, primarily dependent on the features (Liu et al., 2004; Li et al., 2010) at the cost of labour and efficiency.

With the rapid development of techniques, graph-based methods (Linmei et al., 2019; Li et al., 2020; Xu et al., 2020) have been adopted for text summarization, notably using graph structures for better modeling relations between words. Though remarkable performance has been achieved, existing methods attempt to model either *word relations* or *structure information*, instead of model them concurrently. For example, (Xu et al., 2020) models *structure information* between sub-sentences obtained from Rhetorical Structure Theory (RST) (Mann and Thompson, 1988). (Zhu et al., 2020) constructs a knowledge graph which captures *word relations*, by extracting triples, i.e. (subject, relation, object) from text.

In this paper, we propose DG-ABS, a Dual Graph Neural network for Abstractive Sentence Summarization, to *simultaneously model word relations and structure information* from given sentences. In particular, we leverage FrameNet (Fillmore et al., 1976; Baker et al., 1998), a semantic database that provides schematic scenario representation, to construct *Semantic Scenario Graph* (SSG) and *Semantic Word Relation Graph* (SWRG).

In FrameNet, *Frame* (F) is defined as a composition of *Lexical Units* (LUs) and a set of *Frame Elements* (FEs). Given a sentence, if its certain word evokes a Frame by matching a LU, then it is called *Target* (T) (Guo et al., 2020). Taking Frame **Leadership** in Figure 1 as an example, the word *premier* evokes the Frame, which contains two FEs, i.e., *Governed*, *Leader*. The FE *Governed* is *filled* by word *Chinese*. It is worth mentioning that FrameNet connects different relevant Frames into *a Frame network* by defining **Frame-to-Frame** (F-to-F) relations, which provide natural and effective ways to model *semantic relations*. The connected Frames, used to build SSG, provide the semantic scenario information at a higher conceptual level. On the other hand, the word relations are used to build SWRG based on *filler words* to Frames and Frame Elements at lower word level. Over these two semantic graphs, Gated Graph Neural Networks (GGNN) (Li et al., 2016) is first built

*Corresponding author: Ru Li.

Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 2522–2529
November 7–11, 2021. ©2021 Association for Computational Linguistics

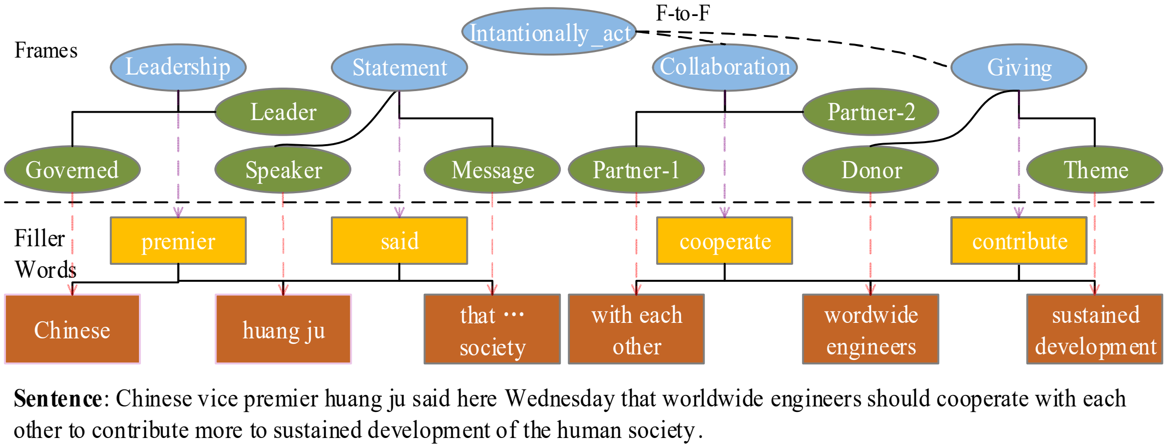

Figure 1: FrameNet-style parsing of the sentence. The top block contains Frame (blue) and Frame Elements (green). The bottom block is the filler words of the Frame (yellow) and Frame Element (brown). The Frame or Frame Element corresponds to its filler words in vertical direction, e.g., Frame Element *Leader* and *Speaker* have the same filler words *huang ju* (red dashed line), *permier* is the Target of Frame **Leadership** (purple dashed line).

to capture word relations and structures individually. Then an attention fusion method is designed to integrate dual graph representations, which will be fed into decoder to generate accurate summary. The contribution of this paper is three-fold:

1. We propose a novel DG-ABS model, which, to the best of our knowledge, is the first attempt to simultaneously capture word relations and structure information to guide the summary generation.
2. We design a graph fusion module between *semantic scenario graph* and *semantic word relation graph*, which further facilitates learning better graph semantic representation.
3. Experimental results show our DG-ABS model achieves competitive performance comparing with state-of-the-art approaches on benchmark Gigaword and DUC 2004 datasets.

2 Methodology

In this section, we introduce the overall architecture of DG-ABS which is shown in Figure 2, consisting of five key modules:

(1) **Graph Construction** aims to obtain the dual graph graph $\mathcal{G}$ (SWRG graph $\mathcal{G}_w$ and SSG graph $\mathcal{G}_s$) by leveraging knowledge from FrameNet.

(2) **Encoder** encodes the given sentence $\mathcal{X}$ and dual graph $\mathcal{G}$ representation respectively.

(3) **Graph Fusion Module** enhances the correlation between SWRG representation h^w and SSG representation h^s to get better semantic representation.

(4) **Feature Aggregation** integrates the graph representation h^g and sentence representation $\mathcal{C}^b$ into an overall semantic representation $\mathcal{C}$.

(5) **Summary Generation** employs the overall representation $\mathcal{C}$ to generate its summary.

2.1 Graph Construction for SSG and SWRG

Semantic Scenario Graph (SSG). As illustrated in Figure 1, the sentence contains four Frames, and each Frame has a set of Frame Elements. Different Frames are connected by F-to-F relations. Each Frame is a nucleus node and more central and important, while each Frame Element is a satellite node, more peripheral and less important in terms of content and grammatical reliance (Xu et al., 2020).

Semantic Word Relation Graph (SWRG). Frame is an abstract semantic scenario, and the same Frame in a sentence may have different filler words. In order to model more fine-grained semantic relations between words, we use the words which filled to the corresponding Frames and Frame Elements to build SWRG. Same as the SSG, each Target word of Frame is a nucleus node and the filler words of Frame Elements are satellite nodes. Then, we use the same F-to-F relations to connect the nucleus nodes to a whole network.

Note, for both SSG and SWRG, an overall sentence level (root) node will be added to connect nodes together, if the nucleus nodes can not be connected to a whole network by F-to-F relations.

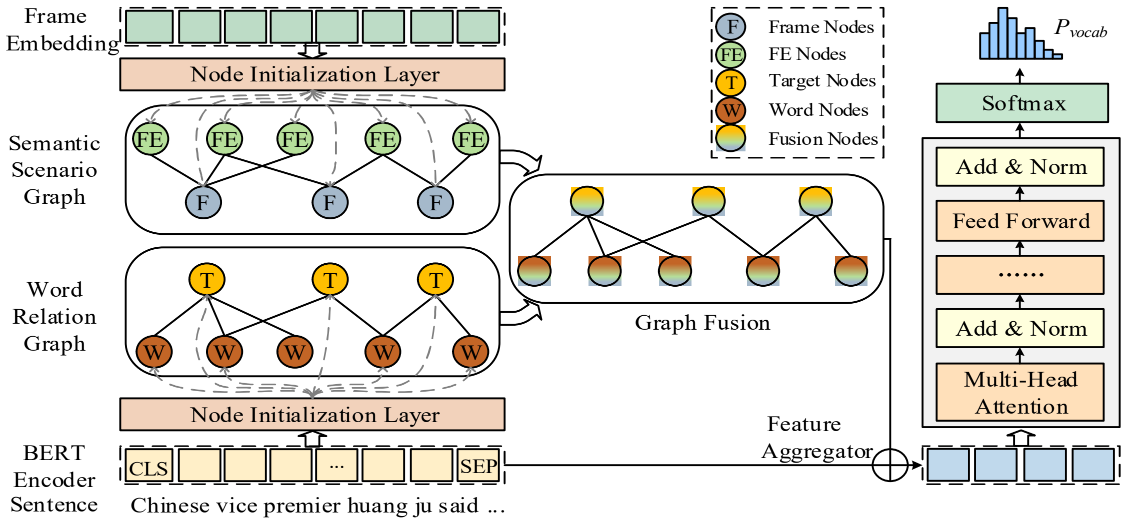

Figure 2: Model architecture of DG-ABS.

2.2 Sentence and Graph Encoder

Sentence Encoder. Given the source sentence $\mathcal{X} = [w_1, w_2, ..., w_i, ..., w_l]$, where w_i is the i-th word in sentence, and l is the length of the sentence. We employ the pre-trained BERT (Devlin et al., 2019) to construct its contextual information for each token, and produce a sequence of contextual representation $\mathcal{C}^b = [x_1, x_2, ..., x_i, ..., x_l]$, where x_i is the i-th corresponding hidden state.

$$\mathcal{C}^b = BERT(\mathcal{X}) \tag{1}$$

Graph Encoder. We define the SSG graph $\mathcal{G}_s$ and SWRG graph $\mathcal{G}_w$ in a unified way. Given the constructed SWRG $\mathcal{G}_w = (\mathcal{V}, \mathcal{E})$ as an example, where $\mathcal{V} = \{v_i\}_{i=1:N^v}$, and $\mathcal{E} = \{e_k\}_{k=1:N^e}$ represent graph nodes and the edges, N^v and N^e are the number of nodes and edges, respectively. Each node $v = \{w_j\}_{j=m_v}^{n_v}$ is a text span in $\mathcal{X}$, where m_v/n_v is starting/ending position of the text span. Then, initial representation h_v^0 for each node $v = \{w_j\}_{j=m_v}^{n_v}$ is computed by averaging corresponding text span representations in $\mathcal{C}^b$:

$$h_v^0 = \frac{1}{|m_v - n_v|} \sum_{j=m_v}^{n_v} x_j \tag{2}$$

Note in SSG $\mathcal{G}_s$, each node is a Frame or Frame Element, and the initial representation h_v^0 for each node is initialized by BERT. It encodes the Frame name definition and Frame element definition, and then use the first input token ([CLS]) representation of the last layer as their embeddings respectively.

After building the graph $\mathcal{G}_w$, to dynamically capture the correlation between nodes, we design a K layers Dual Graph Encoder (DGE) which builds upon Gated Graph Neural Network (GGNN) (Li et al., 2016; Pan et al., 2020) to update all the node representations. The input for the k-th DGE layer is the output of the previous layer, denoted as $h^{(k-1)} = \{h_1^{(k-1)}, h_2^{(k-1)}, ..., h_{N^v}^{(k-1)}\}$. The k-th layer state transition $h^{(k)}$ for each node $v \in \mathcal{V}$ can be calculated as follows:

$$h_i^k = \rho\left(\sum_{r\in\mathcal{R}} \sum_{v_j\in\mathcal{N}_i} \alpha_{ij}^{(k-1)} W_r h_j^{(k-1)}\right) \tag{3}$$

where h_i^k is the hidden state of node v_i at k-th DGE layer. $\mathcal{R}$ is the set of edge types, and $\mathcal{N}_i$ is the neighbors of node v_i. W_r denotes learnable parameters, and $\rho(\cdot)$ is an activation function. $\alpha_{ij}^{(k-1)}$ is the attention weight of node v_i over v_j.

$$\alpha_{ij}^{(k-1)} = \frac{exp(h_i^{(k-1)} \cdot h_j^{(k-1)})}{\sum_{j'\in\mathcal{N}_i} exp(h_i^{(k-1)} \cdot h_{j'}^{(k-1)})} \tag{4}$$

After K layers of graph propagation, we obtain final graph $\mathcal{G}_w$ representation $h^w = h^k$. Then, graph $\mathcal{G}_s$ representation h^s can be computed similarly.

2.3 Graph Fusion Module

After the individual graph encoding, branch h^s and h^w are expected to capture the structure-related and word-related features respectively. Following the human writing behavior, they always organize the structure of article first, and then write the article content according to the article structure. Thus, we further fuse the graph representation from the two branches dynamically (Hu et al., 2017; Yin et al., 2020). Specifically, we first calculate the enriched $\mathcal{G}_s$ representation $\hat{h^g}$ as follows:

$$\hat{h_t^g} = \sum_{j \in \mathcal{N}^v} \alpha_{tj} \cdot h_j^w \tag{5}$$

$$\alpha_{t,j} = \sigma(W_1 h_t^s + W_2 h_j^w) \tag{6}$$

Where σ is an activation function. $\mathcal{N}^v$ is the number of nodes, and W_1^l and W_2^l are parameter matrices. Likewise, we can obtain the enriched $\mathcal{G}_w$ representation $\check{h^g}$.

In the second step, we compute an update gate u to fuse the enriched representations. Concretely, we generate the final graph representation h^g in the following way:

$$u_t = \beta(\hat{h_t^g}, \check{h_t^g}) \tag{7}$$

$$h_t^g = (1 - u_t)\hat{h_t^g} + u_t \check{h_t^g} \tag{8}$$

Where β stands for a nonlinear function.

2.4 Feature Aggregation

To obtain the final enriched sentence representation $\mathcal{C}$, we integrate sentence representation $\mathcal{C}^b$ with graph representation h^g. For a word w_i in $\mathcal{X}$, we search all the nodes which contain w_i, denoted as $\{h_1^g, h_2^g, ..., h_k^g\}$. Then, we concatenate the word representation x_i with the average of corresponding node representations for summary generation.

$$\mathcal{C} = [x_i; \frac{1}{k} \sum_{j=1}^{k} h_j^g] \tag{9}$$

Finally, we obtain the enriched sentence representation $\mathcal{C}$ for summary generation.

2.5 Summary Generation

We build a transformer-based decoder (Vaswani et al., 2017; Zhu et al., 2020), which takes the enriched sentence representation $\mathcal{C}$ to generate summary one word at a time. At each decoding step t, the current decoding state s_t is updated by the previous output $[y_1, y_2, ..., y_{t-1}]$ and $\mathcal{C}$. The probability $\mathcal{P}_{(y_t)}$ of next token y_t is represented as:

$$\mathcal{P}_{(y_t)} = softmax\,(W s_t + b) \tag{10}$$

where W and b are learnable parameters.

Method	R-1	R-2	R-L
ProphetNet	39.55	20.27	36.57
ERNIE-GEN	39.25	20.25	36.53
BERTShare	38.13	19.81	35.62
Open-NMT	36.73	17.86	33.68
Re3Sum	37.04	19.03	34.46
BiSET	39.11	19.78	36.87
DG-ABS	41.94	23.58	38.97

Table 1: F-measures ROUGE socres on Gigaword.

3 Experiments

3.1 Data and Evaluation Metrics

We test our proposed framework on the popular dataset Gigaword (Napoles et al., 2012) following previous work (Wang et al., 2019; Xiao et al., 2020). The training, validation, and test set sizes are 3.8M, 189k and 1951 respectively. Additionally, we also apply our framework on the DUC 2004 summarization task (Over et al., 2007). As it only contains 500 news articles, we directly use the model trained on the Gigaword to test on the DUC 2004 dataset which can also evaluate models' generalization capabilities.

We employ standard ROUGE metrics (Lin and Hovy, 2003), including ROUGE-1 (R-1), 2 (R-2), and L (R-L) to evaluate all the models. Following the existing work, we apply recall-based ROUGE metric on DUC 2004 data (Rush et al., 2015; Gao et al., 2019), and F-based ROUGE to evaluate Gigaword data (Cao et al., 2018; Xiao et al., 2020).

3.2 Baselines

We compare with six models for Gigaword data, including three *non-pre-trained models*: **Open-NMT** (Klein et al., 2017), **Re3Sum** (Cao et al., 2018), **BiSET** (Wang et al., 2019), and three *pre-trained models*: **BERTShare** (Rothe et al., 2020), **ProphetNet** (Yan et al., 2020), **ERNIE-GEN** (Xiao et al., 2020). For DUC 2004, we compare with four models, including **Featseq2seq** (Nallapati et al., 2016), **SEASS** (Zhou et al., 2017), **ERAML** (Li et al., 2018), and **GLEAM** (Gao et al., 2019).

3.3 Performance Comparison

Results on Gigaword. As shown in Table 1. We observe that DG-ABS model achieves 41.94, 23.58 and 38.97 in terms of three evaluation metrics, which are 2.83, 3.8, and 2.1 point better than

Method	R-1	R-2	R-L
Featseq2seq	28.61	9.42	25.24
SEASS	29.21	9.56	25.51
ERAML	29.33	10.24	25.24
GLEAM	29.51	9.78	25.60
DG-ABS	30.03	10.71	26.05

Table 2: R-measures ROUGE socres on DUC 2004.

Method	R-1	R-2	R-L
DG-ABS	41.94	23.58	38.97
-w/o SSG	41.02	22.67	38.01
-w/o SWRG	39.29	21.14	36.54
-w/o Dual Graph	36.74	19.83	35.17

Table 3: Ablation study on Gigaword data.

second-best results (from different methods), indicating we are able to generate better quality summaries. In addition, compared with ProphetNet (pre-trained model) and BiSET (non-pre-trained model) which employ selective gate and adopt summary templates, our DG-ABS model outperforms them significantly, signifying the importance of leveraging the Frame semantic information.

Results on DUC 2004. Table 2 shows that our model once again achieves the best performance across all three metrics consistently. From the overall results on Gigaword and DUC 2004, we can see that DG-ABS is effective by leveraging fine-grained Frame semantic information into the graph to guide summary generation.

Ablation Study. We also conduct ablation study to assess the impact of different components of DG-ABS. As shown in Table 3, By removing either SSG or SWRG , the performance degrades significantly, indicating both SSG and SWRG are important to our overall model. When we do not use our dual graph at all, the performance degrades most, verifying these two innovative steps play crucial roles for generating high quality summaries. Detailed case analysis is available in appendix A.

4 Conclusion

We propose a novel DG-ABS model to *simultaneously* capture the word relations and structure information from sentences to effectively guide summary generation. Specifically, we first build *semantic word relation graph* and *semantic scenario graph* based on FrameNet, and subsequently design a graph fusion method to enhance their correlation and enriched joint representations. Extensive experimental results on two popular datasets demonstrate our model achieves better performance than state-of-the-art approach. In future work, we would like to do several experiments on other related tasks to test the versatility of our framework. Also, we plan to use the semantic information from FrameNet to investigate the problem of summarization evaluation.

Acknowledgements

We would like to thank all the anonymous reviewers for providing valuable feedback. This work was sponsored by the National Key Research and Development Program of China under grant (2020AAA0106100) and National Natural Science Foundation of China (No.61936012, No.62176145).

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics.

Ziqiang Cao, Wenjie Li, Sujian Li, and Furu Wei. 2018. Retrieve, rerank and rewrite: Soft template based neural summarization. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 152–161, Melbourne, Australia. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Charles J Fillmore et al. 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, volume 280, pages 20–32.

Yang Gao, Yang Wang, Luyang Liu, Yidi Guo, and Heyan Huang. 2019. Neural abstractive summarization fusing by global generative topics. *Neural Computing and Applications*, pages 1–10.

Shaoru Guo, Ru Li, Hongye Tan, Xiaoli Li, Yong Guan, Hongyan Zhao, and Yueping Zhang. 2020. A frame-based sentence representation for machine reading comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 891–896, Online. Association for Computational Linguistics.

Linmei Hu, Juan-Zi Li, Liqiang Nie, Xiaoli Li, and Chao Shao. 2017. What happens next? future subevent prediction using contextual hierarchical l-stm. In *AAAI*.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M Rush. 2017. Opennmt: Open-source toolkit for neural machine translation. *arXiv preprint arXiv:1701.02810*.

Kevin Knight and Daniel Marcu. 2002. Summarization beyond sentence extraction: A probabilistic approach to sentence compression. *Artificial Intelligence*, 139(1):91–107.

Haoran Li, Junnan Zhu, Jiajun Zhang, and Chengqing Zong. 2018. Ensure the correctness of the summary: Incorporate entailment knowledge into abstractive sentence summarization. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1430–1441, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Wei Li, Xinyan Xiao, Jiachen Liu, Hua Wu, Haifeng Wang, and Junping Du. 2020. Leveraging graph to improve abstractive multi-document summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6232–6243, Online. Association for Computational Linguistics.

Xiao-Li Li, Lei Zhang, Bing Liu, and See-Kiong Ng. 2010. Distributional similarity vs. pu learning for entity set expansion. In *Proceedings of the ACL 2010 Conference Short Papers*, ACLShort '10, page 359–364, USA. Association for Computational Linguistics.

Yujia Li, Richard Zemel, Marc Brockschmidt, and Daniel Tarlow. 2016. Gated graph sequence neural networks. In *Proceedings of ICLR'16*.

Chin-Yew Lin and Eduard Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 150–157.

Hu Linmei, Tianchi Yang, Chuan Shi, Houye Ji, and Xiaoli Li. 2019. Heterogeneous graph attention networks for semi-supervised short text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4821–4830, Hong Kong, China. Association for Computational Linguistics.

Bing Liu, Xiaoli Li, Wee Sun Lee, and Philip S. Yu. 2004. Text classification by labeling words. In *Proceedings of the 19th National Conference on Artifical Intelligence*, AAAI'04, page 425–430. AAAI Press.

William C Mann and Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Ramesh Nallapati, Bowen Zhou, Caglar Gulcehre, Bing Xiang, et al. 2016. Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv:1602.06023*.

Courtney Napoles, Matthew Gormley, and Benjamin Van Durme. 2012. Annotated gigaword. In *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction*, pages 95–100. Association for Computational Linguistics.

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

Liangming Pan, Yuxi Xie, Yansong Feng, Tat-Seng Chua, and Min-Yen Kan. 2020. Semantic graphs for generating deep questions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1463–1475, Online. Association for Computational Linguistics.

Sascha Rothe, Shashi Narayan, and Aliaksei Severyn. 2020. Leveraging pre-trained checkpoints for sequence generation tasks. *Transactions of the Association for Computational Linguistics*, 8:264–280.

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, Lisbon, Portugal. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, undefinedukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA. Curran Associates Inc.

Kai Wang, Xiaojun Quan, and Rui Wang. 2019. BiSET: Bi-directional selective encoding with template for abstractive summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2153–2162, Florence, Italy. Association for Computational Linguistics.

Dongling Xiao, Han Zhang, Yukun Li, Yu Sun, Hao Tian, Hua Wu, and Haifeng Wang. 2020. Ernie-gen: An enhanced multi-flow pre-training and fine-tuning framework for natural language generation.

Jiacheng Xu, Zhe Gan, Yu Cheng, and Jingjing Liu. 2020. Discourse-aware neural extractive text summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5021–5031, Online. Association for Computational Linguistics.

Yu Yan, Weizhen Qi, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang, and Ming Zhou. 2020. Prophetnet: Predicting future n-gram for sequence-to-sequence pre-training. *arXiv preprint arXiv:2001.04063*.

Yongjing Yin, Fandong Meng, Jinsong Su, Chulun Zhou, Zhengyuan Yang, Jie Zhou, and Jiebo Luo. 2020. A novel graph-based multi-modal fusion encoder for neural machine translation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3025–3035, Online. Association for Computational Linguistics.

Liang Zhou and Eduard Hovy. 2004. Template-filtered headline summarization. In *Text Summarization Branches Out*, pages 56–60.

Qingyu Zhou, Nan Yang, Furu Wei, and Ming Zhou. 2017. Selective encoding for abstractive sentence summarization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1095–1104, Vancouver, Canada. Association for Computational Linguistics.

Chenguang Zhu, William Hinthorn, Ruochen Xu, Qingkai Zeng, Michael Zeng, Xuedong Huang, and Meng Jiang. 2020. Boosting factual correctness of abstractive summarization with knowledge graph. *arXiv 2003.08612*.

A Additional Cases Analysis

Recall that SWRG and SSG are two key components of our DG-ABS model. To determine their individual effects, We also have analyzed many cases on Gigaword dataset, some of them are presented in Table 4. Our DG-ABS model performs well in generating an accurate and informative summary as well as integrating Frame semantic information. Take the first case in Table 4 as an example, without Dual Graph (-w/o Dual Graph), the system does not generate key topic *sparked a violent reaction*. Compare to -w/o SSG or -w/o SWRG, DG-ABS model captures more complete key information and generates new word to increase diversity. We will share the implementations if the paper gets accepted in future time.

Case 1	
Source Sentence	Indian prime minister p.v. Narasimha Rao ' s promise of more autonomy for troubled Kashmir and his plea for early state elections has sparked a violent reaction from provincial Moslem and opposition parties .
Reference Summary	Indian pm ' s announcement on Kashmir polls autonomy sparks outrage
-w/o Dual Graph	Indian pm ' s promise more autonomy in Kashmir
-w/o SSG	Indian pm ' s promise of autonomy sparks anger
-w/o SWRG	Indian pm ' s call for autonomy for elections sparks violent
DG-ABS	Indian pm ' s pledge of autonomy for Kashmir sparks violent reaction
Case 2	
Source Sentence	Japan ' s toyota team europe were banned from the world rally championship for one year here on friday in a crushing ruling by the world council of the international automobile federation .
Reference Summary	Toyota are banned for a year
-w/o Dual Graph	Israel prepares for rabin ' s state funeral
-w/o SSG	Toyota europe banned for one year
-w/o SWRG	Toyota banned from rally championship for # year
DG-ABS	Toyota europe are banned for a year
Case 3	
Source Sentence	India won the toss and chose to bat on the opening day in the opening test against west indies at the antigua recreation ground on friday .
Reference Summary	India win toss and elect to bat in first test
-w/o Dual Graph	India win toss bat in first test against west indies
-w/o SSG	India wins toss elects to bat in # st test
-w/o SWRG	India wins toss and bat in opening test
DG-ABS	India wins toss and elects to bat in opening test

Table 4: The examples of summaries generated by different models.