Knowledge and Keywords Augmented Abstractive Sentence Summarization

Anonymous EMNLP submission

Abstract

In this paper, we study the knowledge-based abstractive sentence summarization. There are two essential information features that can influence the quality of news summarization, which are topic keywords and the knowledge structure of the news text. Besides, the existing knowledge-augmented methods have poor performance on sentence summarization since the sparse knowledge structure problem. Considering these, we propose **KAS**, a novel Knowledge and Keywords Augmented Abstractive Sentence Summarization framework. Tri-encoders are utilized to integrate contexts of original text, knowledge structure and keywords topic simultaneously, with a special linearized knowledge structure. Automatic and human evaluations demonstrate that KAS achieves the best performances.[1]

1 Introduction

With the increasing of computing power and model capacity, it is possible to generate mostly grammatical summarization of natural language text. In general, there are two essential information features of summarization: (1) topic keywords in text (2) the knowledge structure of the text. These features can basically cover all the information in summary generation, especially in sentence or short text summarization. Therefore, considering this reason, we are building a neural network model that integrates both topic keyword context and knowledge structure context.

Knowledge augmented summarization has been intensively studied recently, most of which are about document summarization. However, there is not much research on knowledge-based sentence summarization. The main reason is that the existing methods are not applicable to sentence summarization. The knowledge-based summarization frameworks usually use GNN as the knowledge structure encoder. However, the knowledge graph of sentence is usually sparse, and GNN has poor performance in sparse knowledge structure. Specifically, GNNs may cause over-smoothing problem when training on the sparse graphs(Alon and Yahav, 2021), especially for GCNs (Kipf and Welling, 2017), decreasing the robustness and performance of the model. Therefore, we are creating a new knowledge-augmented sentence summarization model considering these problems. Besides, considering most of the knowledge based summarization models are only applicable to English, we are aiming at making our model applicable to multiple languages.

In order to address these issues, we propose a special linearized knowledge sequence structure that are applicable to sentence summarization. Correspondingly, we propose a novel tri-encoder framework KAS integrating three separate encoders, considering contexts of original text, topic keywords and knowledge structure simultaneously based on their salience. Evaluations demonstrate that KAS framework and the corresponding linearized knowledge structure enhances the performances significantly. Besides, the structure of KAS can be applied to summarization on multiple languages. We have conducted experiments on English and Chinese corpus and achieved best performances on both.

2 Related Work

Knowledge-based Summarization The existing method for utilizing knowledge graph into text generation and summarization is adding a separate encoder to encode the vectorized knowledge graph for context integration. Ribeiro et al. (2020) introduced a knowledge graph encoding strategy for graph-to-text generation model. Koncel-Kedziorski et al. (2019), Huang et al. (2020) proposed a text generation (summarization) model integrated with a GNN encoder (Veličković et al., 2018) using encoded graph data preprocessed from the input text.

[1] Code is in https://github.com/SeanG-325/KAS

Algorithm 1 Knowledge Sequence Construction

Require: Text Sequence $\mathcal{S}$; Triples Extractor $\mathcal{E}$; Knowledge Graph $\mathcal{G}_k[*][*] = 0$.
  $T = \mathcal{E}(\mathcal{S})$
  for all $e \in T$ **do**
    if $\mathcal{E}(e.E) \neq \phi$ **then**
      $e.E = \mathcal{E}(E).E$
      $T = T \cup \mathcal{E}(E)$
    end if
  end for
  for $e \in T$ **do**
    $\mathcal{G}_k[e.E_1][e.R] = \mathcal{G}_k[e.R][e.E_2] = 1$
  end for
  Collect the occurance locations in $\mathcal{S}$ for all vertices in $\mathcal{G}_k$ as $L = \{l_1, ..., l_m\}$
  $S_{KG} = \text{DFS}_{KGL}(\mathcal{G}_k, L)$
  return S_{KG}

Aiming at solving the possible sparse problem of graphs, Konstas et al. (2017) proposed a method of graph linearization, and used LSTM to encode the graph structure.

Pointer Mechanism Pointer Mechanism has drawn much attention in text generation(Miao and Blunsom, 2016; Nallapati et al., 2016; Gulcehre et al., 2016; Eric and Manning, 2017). In text summarization, Pointer-Generator Network model (See et al., 2017) is proposed to keep the generation ability while using pointer mechanism. Sun et al. (2018) proposed a method for using pointer mechanism with multiple separate encoders. The idea of Pointer Mechanism is setting soft or hard gates to select from predefined vocabulary or input sequences to generate tokens.

3 Summarization Framework

3.1 Knowledge and Keywords Construction

The whole linearized knowledge graph constructing process is presented in Algorithm 1. For $\mathcal{E}$, we use OLLIE (Mausam et al., 2012) to extract triples from English news texts. As few established tools are for open domain Chinese triple extraction, we extract triples from semantic rules using Language Technology Platform (Che et al., 2010). As shown in the algorithm, fact triples in different granularity will be re-extracted until the granularity of all triples is minimized, and we keep reconstructing the triples to enhance the connectivity of the knowledge graph. We assume each triple e has 3 elements: E_1, R and E_2. The E in $e.E$ denotes to

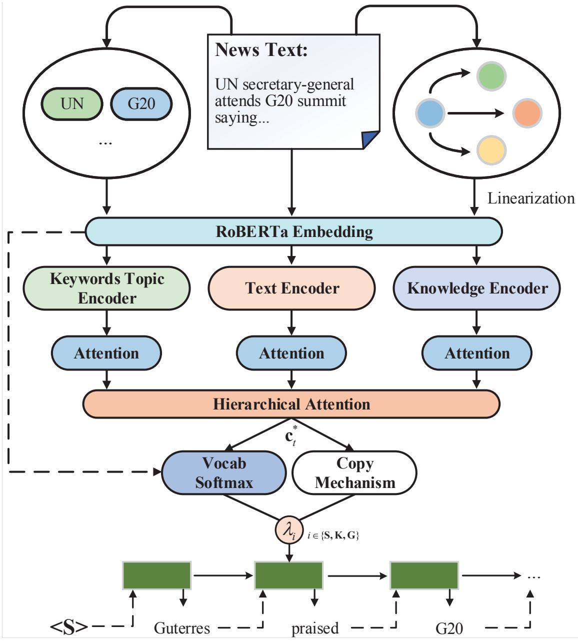

Figure 1: The Model Structure of KAS. The λ_i are soft gates for distributing copy probabilities.

E_1 and E_2. Then all edges (relationships) will be converted to vertices.

We then generate the linearized knowledge graph sequence by a modified DFS algorithm. The DFS is modified on the start vertex selection and priorities of different traversal paths. When traversal starts or the current vertex has more than one path, we select the vertex whose token first appear in the source text as the next. It reduces text redundance effectively and makes the framework focus more on the key logic instead of other irrelevant information.

For keywords topic sequences, we use TextRank(Mihalcea and Tarau, 2004) algorithm to extract keywords from source text, and make them in the order in which they appear in the original text. This brings priori topic knowledge to the model and makes the model explicitly consider the keywords topic information of the text.

3.2 Architecture

KAS takes as input a news text $\boldsymbol{S} = \{x_i\}$, a keywords topic sequence $\boldsymbol{K} = \{k_i\}$ and a knowledge sequence $\boldsymbol{G} = \{v_i\}$, and let $\mathcal{D} = \{\boldsymbol{S}, \boldsymbol{K}, \boldsymbol{G}\}$. The tri-encoder structure shown in Figure 1 integrates the context of original source text, keywords topic and internal knowledge. The RoBERTa (Liu et al., 2019) is utilized for word embedding pre-training, and we use the outputs of the last RoBERTa layer as the input embedding for all encoders. We build encoders to the generate hidden states $\boldsymbol{h}_t^{\boldsymbol{S}}, \boldsymbol{h}_t^{\boldsymbol{K}}, \boldsymbol{h}_t^{\boldsymbol{G}}$, which is $\boldsymbol{h}_t^{\boldsymbol{x}} = g(\boldsymbol{h}_{t-1}^{\boldsymbol{x}})(\boldsymbol{x} \in \mathcal{D})$, in which function g is a bi-directional LSTM. The hidden states in the final time step of the three encoders, $\boldsymbol{h}_{l_1}^{\boldsymbol{S}}$, $\boldsymbol{h}_{l_2}^{\boldsymbol{K}}$, $\boldsymbol{h}_{l_3}^{\boldsymbol{G}}$, should be transformed into the decoder initial state $\boldsymbol{d}_0 = \tanh(\boldsymbol{W}_m \cdot [\boldsymbol{h}_{l_1}^{\boldsymbol{S}} || \boldsymbol{h}_{l_2}^{\boldsymbol{K}} || \boldsymbol{h}_{l_3}^{\boldsymbol{G}}])$.

The attentions of the source text, keywords and knowledge are computed as($\boldsymbol{\alpha}_t^{\boldsymbol{S}}$), ($\boldsymbol{\alpha}_t^{\boldsymbol{K}}$) and ($\boldsymbol{\alpha}_t^{\boldsymbol{G}}$) (Bahdanau et al., 2015). The context vectors are computed as $\boldsymbol{c}_t^{\boldsymbol{x}} = \sum_{i=0}^{t} \alpha_i^{\boldsymbol{x}} \boldsymbol{h}_i^{\boldsymbol{x}}, \quad \boldsymbol{x} \in \mathcal{D}$.

We decode with an attention-based decoder, the decoder hidden state at timestep t $\boldsymbol{d}_t$ is $\boldsymbol{d}_t = f(\boldsymbol{d}_{t-1}, \boldsymbol{c}_{t-1}^{\boldsymbol{S}}, \boldsymbol{c}_{t-1}^{\boldsymbol{K}}, \boldsymbol{c}_{t-1}^{\boldsymbol{G}}, \boldsymbol{y}_{t-1})$, in which $\boldsymbol{d}_{t-1}$ is the decoder hidden state, $\boldsymbol{y}_{t-1}$ is the decoder input, $\boldsymbol{c}_{t-1}^{\boldsymbol{x}}$ are the context vectors. The function f denotes to an unidirectional LSTM.

Hierarchical Attention The salience for the three contexts should be automatically adjusted. Therefore, besides the word-level attention in each encoder, we further utilize a encoder-level hierarchical attention mechanism for ensemble context. We compute the ensemble attention as

$$b^{\boldsymbol{x}} = \boldsymbol{u}^T \tanh(\boldsymbol{W}_{hc}^{\boldsymbol{x}} \boldsymbol{c}_t^{\boldsymbol{x}} + \boldsymbol{W}_{hd}^{\boldsymbol{x}} \boldsymbol{d}_t + \boldsymbol{b}_h^{\boldsymbol{x}})$$

$$\beta^{\boldsymbol{x}} = \mathrm{softmax}(b^{\boldsymbol{x}}), \quad \boldsymbol{x} \in \mathcal{D}$$

$\beta^{\boldsymbol{x}}$ is the hierarchical attention weight of the three contexts in the ensemble context. We then compute the ensemble context $\boldsymbol{c}_t^*$ as

$$\boldsymbol{c}_t^* = \sum_{\boldsymbol{x} \in \mathcal{D}} \beta^{\boldsymbol{x}} \boldsymbol{c}_t^{\boldsymbol{x}}$$

The ensemble context $\boldsymbol{c}_t^*$ is a fixed length vector encoding salient information from the three contexts of the tri-encoder model.

$P_{vocab}(w)$ is calculated by scaling $[\boldsymbol{h}_t || \boldsymbol{c}_t^*]$ to the vocabulary size and taking a softmax:

$$P_{vocab}(w) = \mathrm{softmax}(\boldsymbol{W}_s [\boldsymbol{h}_t || \boldsymbol{c}_t^*] + b_s)$$

To allow $\boldsymbol{W}_s$ to reuse the linguistic in input embedding and decrease the number of parameters, we integrate weight-sharing mechanism (Paulus et al., 2018) in the model as $\boldsymbol{W}_s = \tanh(\boldsymbol{W}_{emb} \cdot \boldsymbol{W}_{sh})$, in which $\boldsymbol{W}_{emb}$ is input embedding matrix.

Tri-Copy Mechanism We compute p_{cpy}, which is overall copy probability and will be distributed to the three encoders:

$$p_{cpy} = \sigma(\boldsymbol{W}_{cpy} [\boldsymbol{h}_t || \boldsymbol{c}_t^*] + \boldsymbol{b}_{cpy})$$

$P_{cpy}(w)$ is distributed to the tri-encoders with soft gates $\lambda_S,\ \lambda_K,\ \lambda_G$. Here, $\lambda_i (i \in \mathcal{D})$ automatically adjust $\boldsymbol{d}_t$, $\boldsymbol{y}_{t-1}$, and the context vector $\boldsymbol{c}_t^i$. We define λ_i as:

$$\lambda_i = \frac{\sigma(\boldsymbol{w}_{di}^T \boldsymbol{d}_t + \boldsymbol{w}_{yi}^T \boldsymbol{y}_{t-1} + \boldsymbol{w}_{ci}^T \boldsymbol{c}_t^i)}{\sum_{\boldsymbol{x}} \sigma(\boldsymbol{w}_{d\boldsymbol{x}}^T \boldsymbol{d}_t + \boldsymbol{w}_{y\boldsymbol{x}}^T \boldsymbol{y}_{t-1} + \boldsymbol{w}_{c\boldsymbol{x}}^T \boldsymbol{c}_t^{\boldsymbol{x}})} \cdot p_{cpy} \qquad (i, \boldsymbol{x} \in \mathcal{D})$$

The training loss can be defined as the the negative log likelihood of the target sequence:

$$\mathcal{L} = -\sum_{t=0}^{T} \log p(y_t = w_t^* | P_{vocab}, \boldsymbol{S}, \boldsymbol{K}, \boldsymbol{G}, y_{<t})$$

in which w_t^* is the target word at step t, T is the length of the target sequence. The multi-copy mechanism can make the model more inclined to generate informative words.

4 Experiments

4.1 Dataset

We use LCSTS dataset (Hu et al., 2015), which contains a training set of 2.4M online Chinese short news texts in Chinese social media SinaWeibo. We choose 725 pairs from the test set with high annotation scores as our test set. Besides, we consider the annotated Gigaword corpus (Rush et al., 2015), which leads to around 3.8M training samples and 1951 test samples for evaluation.

4.2 Experiment Settings

The model is mainly implemented in Tensorflow[2]. In the data preprocess step, we use Jieba[3] for Chinese words segmentation and topic keywords extraction, and LTP(Che et al., 2010) for knowledge extraction. For English we use OLLIE to extract knowledge triples. For our model, we have 512-dimensional hidden states and word embedding. We use a predefined vocabulary of 60k words for both source and target in word-level inputs. Adagrad optimizer is used with learning rate 0.15 and an initial accumulator value of 0.1. All models are trained on a single NVIDIA RTX 2080 Ti GPU, with a batch size of 64 on inputs.

4.3 Automatic and Human Evaluation

The model is evaluated with the standard ROUGE metric (Lin, 2004), shown in Table 1 and 2. We use the F_1 scores for ROUGE-1, ROUGE-2 and ROUGE-L.

Models	RG-1	RG-2	RG-L
PGEN+COV(See et al., 2017)	38.22	25.80	35.46
GLOBAL(Lin et al., 2018)	39.40	26.90	36.50
NCLS(Zhu et al., 2019)	39.71	27.45	37.13
CATT(Duan et al., 2019)	44.35	30.65	40.58
LEXICON(Wan et al., 2020)	42.3	29.8	38.4
KAS			
+Kw	40.74	27.30	36.96
+KG	43.04	30.01	38.82
+KwKG	**44.42**	**31.07**	**40.71**

Table 1: F_1 scores on the LCSTS dataset in terms of the full-length RG-1, RG-2, and RG-L. **Bold** means the best. "+KG" and "+Kw" means the model augmented by knowledge and keywords separately.

Models	RG-1	RG-2	RG-L
SEASS(Zhou et al., 2017)	36.15	17.54	33.63
GLOBAL(Lin et al., 2018)	36.30	18.00	33.80
GENPARSE(Song et al., 2020)	36.61	18.85	34.33
CPDS(Wang et al., 2019)	37.01	17.10	34.87
KAS			
+Kw	36.74	17.51	33.73
+KG	37.01	18.02	34.47
+KwKG	**37.46**	**18.89**	**35.01**

Table 2: F_1 scores on the Gigaword dataset in terms of the full-length RG-1, RG-2, and RG-L. **Bold** means the best. *Italics* means it close to the best score.

Besides the automatic evaluation, we further conduct human evaluation for the framework. We randomly sample 100 articles from LCSTS test set and ask 3 Chinese native speakers to rate summaries of our systems and the baseline (PGEN+COV), along with outputs by human-written summaries. After reading the articles, each judge scores summaries on a Likert scale from 1 (worst) to 5 (best) on (1)*informativeness* and (2)*fluency*. Besides, in the experiment we noticed that the outputs of KAS are more diversified and attractive to readers, so we test (3)*diversity*: whether the summary arouses annotators' reading interest. We consider two types of unfaithful errors: (i) *hallucination error* and (ii) *logical error*. We ask the annotators to label each type as 1 for existence of errors and 0 otherwise.

4.4 Analysis

The automatic evaluation scores show that KAS achieves bests on LCSTS and Gigaword. The Table 3 shows that KAS augmented by both keywords topic and knowledge achieves the best results in almost all indicators, with significant enhancements. We find that diversity of summaries is enhanced

Case Study

ST: 中石化计划推进下游油品销售业务的产权重整，被誉为央企发展混合所有制、打破垄断的一大突破。民企老板直言，如不放开加油站的油源垄断，股权层面出让部分空间的意义有限：“不解决油源，让民资参股只是个花活。”

ST: Sinopec plans to promote the property right reorganization of downstream oil sales business, which is known as a breakthrough for central enterprises to develop mixed ownership and break monopoly. The private enterprise boss said frankly that if the monopoly of oil sources in gas stations is not released, the significance of transferring part of the space at the equity level is limited: "it's meaningless to let private capital participate in the shares without solving the problem of oil source."

Ref: 民营油企负责人：不解决油源让民资参股只是个花活

Ref: Private oil enterprises boss: it's meaningless to let private capital participate in shares without solving the problem of oil sources

+Kw: 民企老板：让民资参股只是个花活

+Kw: Private oil enterprise boss: it's meaningless to let private capital participate in shares

+KwKG: 民企老板炮轰中石化计划：不解决油源让民资参股只是个花活

+KwKG: Private oil enterprise boss bombards Sinopec plan: it's meaningless to let private capital participate in shares without solving the problem of oil sources.

Figure 2: A case study on the LCSTS dataset. **ST** is source text; **Ref** is reference summary; **+Kw** is keywords augmented; **+KwKG** is keywords and knowledge augmented.

Models	Inf.↑	Flu.↑	Div.↑	HE.↓	LE.↓
PGEN+COV	2.92	3.43	3.52	18%	35%
KAS					
+Kw	2.97	3.92*	3.47	10%	31%
+KwKG	3.82*	4.47*	4.15*	**3%***	12%*
HUMAN	**4.30**	**4.63**	**4.48**	17%	**2%**

Table 3: Human evaluation on informativeness (Inf.), fluency (Flu.) and diversity (Div.) (1-to-5), and hallucination error(HE.) and logical error (LE.) (0-to-1). **Bold** are the bests. $\star$: Significantly different from all other models. ($p < 0.05$)

significantly. A case study on LCSTS is shown in Figure 2. As LCSTS is a dataset of social media news in an eye-catching style, we speculate while the knowledge structure may enhance the understanding ability of the framework, it can implicitly enhance the memory of the styles of the training set.

5 Conclusion

In this work, we propose KAS, an abstractive summarization framework augmented by knowledge and topic keywords that supports multiple languages. Experimental results show that KAS generates more qualified summaries and achieves the best performances. In the future, we aim at enhancing attractiveness of sentence summarization based on our structure.

[2]https://www.tensorflow.org/
[3]https://github.com/fxsjy/jieba

References

Uri Alon and Eran Yahav. 2021. On the bottleneck of graph neural networks and its practical implications. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Wanxiang Che, Zhenghua Li, and Ting Liu. 2010. LTP: A Chinese language technology platform. In *Coling 2010: Demonstrations*, pages 13–16, Beijing, China. Coling 2010 Organizing Committee.

Xiangyu Duan, Hongfei Yu, Mingming Yin, Min Zhang, Weihua Luo, and Yue Zhang. 2019. Contrastive attention mechanism for abstractive sentence summarization. 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 3044–3053, Hong Kong, China. Association for Computational Linguistics.

Mihail Eric and Christopher Manning. 2017. A copy-augmented sequence-to-sequence architecture gives good performance on task-oriented dialogue. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 468–473, Valencia, Spain. Association for Computational Linguistics.

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)*, pages 140–149, Berlin, Germany. Association for Computational Linguistics.

Baotian Hu, Qingcai Chen, and Fangze Zhu. 2015. LCSTS: A large scale Chinese short text summarization dataset. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1967–1972, Lisbon, Portugal. Association for Computational Linguistics.

Luyang Huang, Lingfei Wu, and Lu Wang. 2020. Knowledge graph-augmented abstractive summarization with semantic-driven cloze reward. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5094–5107, Online. Association for Computational Linguistics.

Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the 5th International Conference on Learning Representations*, ICLR '17.

Rik Koncel-Kedziorski, Dhanush Bekal, Yi Luan, Mirella Lapata, and Hannaneh Hajishirzi. 2019. Text Generation from Knowledge Graphs with Graph Transformers. 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 2284–2293, Minneapolis, Minnesota. Association for Computational Linguistics.

Ioannis Konstas, Srinivasan Iyer, Mark Yatskar, Yejin Choi, and Luke Zettlemoyer. 2017. Neural AMR: Sequence-to-sequence models for parsing and generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 146–157, Vancouver, Canada. Association for Computational Linguistics.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Junyang Lin, Xu Sun, Shuming Ma, and Qi Su. 2018. Global encoding for abstractive summarization. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 163–169, Melbourne, Australia. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv e-prints*, arXiv:1907.11692.

Mausam, Michael Schmitz, Stephen Soderland, Robert Bart, and Oren Etzioni. 2012. Open language learning for information extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 523–534, Jeju Island, Korea. Association for Computational Linguistics.

Yishu Miao and Phil Blunsom. 2016. Language as a latent variable: Discrete generative models for sentence compression. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 319–328, Austin, Texas. Association for Computational Linguistics.

Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, Barcelona, Spain. Association for Computational Linguistics.

Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Çağlar Gulçehre, and Bing Xiang. 2016. Abstractive text summarization using sequence-to-sequence RNNs and beyond. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 280–290, Berlin, Germany. Association for Computational Linguistics.

Romain Paulus, Caiming Xiong, and Richard Socher. 2018. A deep reinforced model for abstractive summarization. In *International Conference on Learning Representations*.

Leonardo F. R. Ribeiro, Yue Zhang, Claire Gardent, and Iryna Gurevych. 2020. Modeling global and local node contexts for text generation from knowledge graphs. *Transactions of the Association for Computational Linguistics*, 8:589–604.

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.

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)*, pages 1073–1083, Vancouver, Canada. Association for Computational Linguistics.

Kaiqiang Song, Logan Lebanoff, Qipeng Guo, Xipeng Qiu, Xiangyang Xue, Chen Li, Dong Yu, and Fei Liu. 2020. Joint parsing and generation for abstractive summarization. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 8894–8901. AAAI Press.

Fei Sun, Peng Jiang, Hanxiao Sun, Changhua Pei, Wenwu Ou, and Xiaobo Wang. 2018. Multi-source pointer network for product title summarization. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, pages 7–16, New York, NY, USA. Association for Computing Machinery.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lió, and Yoshua Bengio. 2018. Graph attention networks. In *International Conference on Learning Representations*.

Boyan Wan, Zhuo Tang, and Li Yang. 2020. Lexicon-constrained copying network for chinese abstractive summarization. *CoRR*, abs/2010.08197.

Wenbo Wang, Yang Gao, Heyan Huang, and Yuxiang Zhou. 2019. Concept pointer network for abstractive summarization. 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 3076–3085, Hong Kong, China. Association for Computational Linguistics.

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.

Junnan Zhu, Qian Wang, Yining Wang, Yu Zhou, Jiajun Zhang, Shaonan Wang, and Chengqing Zong. 2019. NCLS: Neural cross-lingual summarization. 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 3054–3064, Hong Kong, China. Association for Computational Linguistics.

Appendices

A Details of Human Evaluation

Here we show the details of the indicators in human evaluation.

Informativeness It is the indicator reflecting whether the generated summary covers all important information points in the input text.

Fluency The indicator reflecting whether the summary is grammatically correct, clear and coherent.

Diversity The indicator reflecting whether the summary arouses annotators' reading interest(which is a key quality indicator of social media news summaries).

Logical Error The error for model of generating summaries whose logic structures contradicting with which in the original text (such as summarizing "A is B's dog" as "B is A's dog").

Hallucination Error The error for model of generating summaries containing the facts that are not in or cannot be inferred from original text.

B Case Study

For details and case study, we randomly pick an example of generated summaries in Figure 2. The original examples (in Chinese) are shown and all the texts are carefully translated into English for reading convenience. The words marked in green are key information points in original text, and the words marked in blue are diversified phrase. The examples demonstrate that the combination of knowledge graphs and keywords sequence can increase logicality and diversity in Chinese summarization tasks.

Examples of summary

ST: 中石化计划推进下游油品销售业务的产权重整，被誉为央企发展混合所有制、打破垄断的一大突破。民企老板直言，如不放开加油站的油源垄断，股权层面出让部分空间的意义有限：“不解决油源，让民资参股只是个花活。”
ST: Sinopec plans to promote the property right reorganization of downstream oil sales business, which is known as a breakthrough for central enterprises to develop mixed ownership and break monopoly. The private enterprise boss said frankly that if the monopoly of oil sources in gas stations is not released, the significance of transferring part of the space at the equity level is limited: "it’s meaningless to let private capital participate in the shares without solving the problem of oil source."
Ref: 民营油企负责人：不解决油源让民资参股只是个花活
Ref: Private oil enterprises boss: it's meaningless to let private capital participate in shares without solving the problem of oil sources
+Kw: 民企老板：让民资参股只是个花活
+Kw: Private oil enterprise boss: it's meaningless to let private capital participate in shares
+KwKG: 民企老板炮轰中石化计划：不解决油源让民资参股只是个花活
+KwKG: Private oil enterprise boss bombards Sinopec plan: it's meaningless to let private capital participate in shares without solving the problem of oil sources

ST: 教育部要求每所学校、幼儿园都要制订防止餐桌浪费的具体办法，提倡小份多次管饱的文明用餐方式。各地中小学还要开展餐饮消费、办公用纸、家庭用水等情况的社会调查，到节粮、节水、环保等方面的社会实践基地参与体验活动。
ST: The Ministry of education requires schools and kindergartens to formulate specific measures to prevent table waste, and to promote the civilized way of eating with small portions and full meals for many times. Primary and secondary schools around the country also need to carry out social surveys on catering consumption, office paper and household water consumption, and participate in experience activities in social practice bases of grain saving, water saving and environmental protection.
Ref: 关于在中小学幼儿园广泛深入开展节约教育的意见
Ref: Opinions on extensive and in-depth development of thrift education in primary and secondary school kindergartens
+Kw: 教育部：学校要制订防止餐桌浪费的具体办法
+Kw: Ministry of Education: schools should formulate specific measures to prevent table waste
+KwKG: 教育部要求:学校引导学生勤俭节约
+KwKG: Requirements of the Ministry of Education: schools guide students to be diligent and thrifty

Figure 3: An example of generated summaries on the LCSTS dataset. **ST** is source text; **Ref** is reference summary; **+Kw** is keywords topic augmented; **+KwKG** is keywords topic and knowledge augmented.