

# On the Faithfulness for E-commerce Product Summarization

Peng Yuan, Haoran Li\*, Song Xu, Youzheng Wu, Xiaodong He and Bowen Zhou

JD AI Research

{yuanpeng29, lihaoran24, xusong28}@jd.com  
{wuyouzheng1, xiaodong.he, bowen.zhou}@jd.com

## Abstract

In this work, we present a model to generate e-commerce product summaries. The consistency between the generated summary and the product attributes is an essential criterion for the e-commerce product summarization task. To enhance the consistency, first, we encode the product attribute table to guide the process of summary generation. Second, we identify the attribute words from the vocabulary, and we constrain these attribute words can be presented in the summaries only through copying from the source, i.e., the attribute words not in the source cannot be generated. We construct a Chinese e-commerce product summarization dataset, and the experimental results on this dataset demonstrate that our models significantly improve the faithfulness.

## 1 Introduction

The fast growing of the e-commerce market makes information overload, which hinders users finding the products they need and slows e-commerce platform upgrading the marketing policies. Product summarization technique facilitates user and e-commerce platform with text that contains the most valuable information about products, which is of practical value to address the problem of information overload. In the e-commerce scenarios, unfaithful product summaries that are inconsistent with the corresponding product attributes, e.g., generating “cotton” for a “silk dress”, mislead the users and decrease public credibility of the e-commerce platform. Thus, the faithfulness is a basic requirement for product summarization.

Recently, sequence-to-sequence (seq2seq) methods show promising performances for general text summarization tasks (Rush et al., 2015; Chopra et al., 2016; Zhou et al., 2017; Li et al., 2018; Zhang et al., 2018; Li et al., 2020b), which has been adopted to text generation tasks in the field of e-commerce (Khatri et al., 2018; Yang et al., 2018; Daultani et al., 2019; Chen et al., 2019; Li et al., 2020a). Although applicable, they do not attempt to improve the faithfulness for product summarization. In this paper, we aim to produce faithful product summaries with heterogeneous data, i.e., textual product description and product attribute table, as shown in Figure 1.

First, as the words in product attribute tables are explicit indicators for the product attributes, we propose a dual-copy mechanism that can selectively copy the tokens in textual product descriptions and product attribute words into the summaries. Second, for the product attribute words, we only allow them appearing in the summaries through copying from the source. In this way, the attribute words not belonging to a certain product cannot be presented in the corresponding summary. Thus, the generated summary will not contain incorrect attributes that contradict the product.

Our main contributions are as follows:

- We propose a e-commerce product summarizer that can copy the tokens both from textual product descriptions and the product attribute table.
- We design an attribute-word-aware decoder that guarantees the attribute words can be presented in the summaries only through copying from the source.

---

\*Corresponding author.

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>.

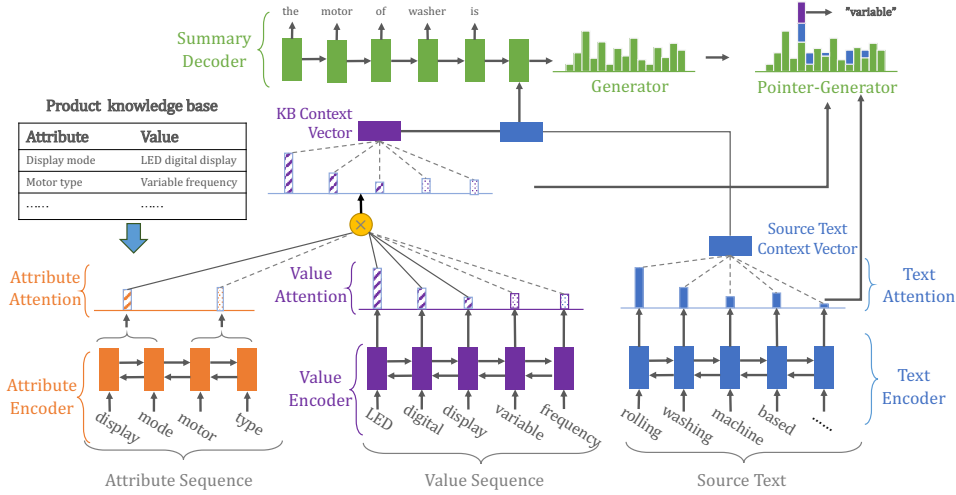


Figure 1: The framework of our model.

- We construct a Chinese e-commerce product summarization dataset that contains approximately half a million product descriptions paired with summaries, and the experimental results on this dataset demonstrate the effectiveness of our model.

## 2 Model

### 2.1 Overview

We start by defining the e-commerce product summarization task. The input is a textual product description and a product attribute table, and the output is a product summary.

As shown in Figure 1, a product attribute table, a.k.a. product knowledge base (Shang et al., 2019), contains a wealth of attribute information for a product. To explore the guidance effect of the product attribute table for producing faithful attribute words, we propose a dual-copy mechanism that can selectively copy tokens from both the textual product description and the product attribute table. To guarantee the unfaithful attribute words not be presented in the summary, all the attribute words are only allowed to appear in the summary through copying.

### 2.2 Dual-Copy Mechanism

Given (1) a product description  $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$  where  $x_i$  is the  $i$ -th word, (2) product attribute set  $\{\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_M\}$  where  $\mathbf{k}_i$  is the  $i$ -th attribute and  $\mathbf{k}_i = \{k_{i,1}, k_{i,2}, \dots, k_{i,K}\}$  where  $k_{i,j}$  is the  $j$ -th word for  $\mathbf{k}_i$ , and (3)  $\mathbf{k}_i$ 's attribute words  $\mathbf{v}_i = \{v_{i,1}, v_{i,2}, \dots, v_{i,V}\}$  where  $v_{i,j}$  is the  $j$ -th work for  $\mathbf{v}_i$ . In Figure 1,  $k_{1,1}$ ="Display",  $k_{1,2}$ ="mode",  $k_{2,1}$ ="Motor",  $k_{2,2}$ ="type",  $v_{1,1}$ ="LED",  $v_{1,2}$ ="digital",  $v_{1,3}$ ="display",  $v_{2,1}$ ="Variable", and  $v_{2,2}$ ="frequency".

A bidirectional LSTM encoder converts  $\mathbf{x}$ ,  $\mathbf{k}$ , and  $\mathbf{v}$  into hidden sequence  $\mathbf{h}^x$ ,  $\mathbf{h}^k$ , and  $\mathbf{h}^v$ . Then, a unidirectional LSTM decoder generates the hidden sequence  $\mathbf{s}$  as follows:

$$s_t = \text{LSTM}(s_{t-1}, y_{t-1}, c_t^x) \quad (1)$$

where  $s_t$  is a hidden state at timestep  $t$ , and  $c_t^x$  is the source sequence context vector that is generated by attention (Bahdanau et al., 2015) mechanism as follows:

$$\alpha_{t,i}^x = \text{softmax}(u_a^T \tanh(W_a h_i^x + V_a s_{t-1})) \quad (2)$$

$$c_t^x = \sum_i \alpha_{t,i}^x h_i^x \quad (3)$$

where  $\alpha_{t,i}^x$  is the attention for  $i$ -th word in the source at timestep  $t$ . Similarly, we can get the attention over each attribute word  $\alpha_{t,i,j}^v$ . We calculate the attribute attention and attribute context vector as follows:

$$\alpha_{t,i}^k = \text{softmax}(u_a^T \tanh(W_a \sum_j h_{i,j}^k + V_a s_{t-1})) \quad (4)$$

$$c_t^k = \sum_i \alpha_{t,i}^k \sum_j h_{i,j}^k \quad (5)$$

Our model is based on the pointer-generator network (PGNet) (Gu et al., 2016; See et al., 2017) that predicts words based on the probability distributions of the generator and the pointer (Vinyals et al., 2015). The generator produces vocabulary distribution  $P_{gen}$  over a fixed target vocabulary as follows:

$$P_{gen}(w) = \text{softmax}(W_b s_t + V_b c_t^x) \quad (6)$$

The dual-pointer copy the word  $w$  from both the source sequence and attribute table. The copy distribution from the source sequence is obtained by the attention distribution over the source sequence:

$$P_{copy}^x(w) = \sum_{i:x_i=w} \alpha_{t,i} \quad (7)$$

We adopt a coarse-to-fine attention (Liu et al., 2019) to calculate the final copy distribution from attribute word sequence with the guidance of attribute-level semantics.

$$P_{copy}^v(w) = \sum_{i:v_{i,j}=w} \alpha_{t,i,j}^v * \alpha_{t,i}^k \quad (8)$$

The overall copy distribution is a weighted sum of the source sequence copy distribution and attribute word copy distribution:

$$P_{copy}(w) = \gamma_t P_{copy}^x(w) + (1 - \gamma_t) P_{copy}^v(w) \quad (9)$$

$$\gamma_t = \sigma(w_d^T c_t^k + w_d^T c_t^x + u_d^T s_t + v_d^T y_{t-1}) \quad (10)$$

The final distribution combines  $P_{gen}$  and  $P_{copy}$ :

$$P(w) = \lambda_t P_{gen}(w) + (1 - \lambda_t) P_{copy}(w) \quad (11)$$

where  $\lambda_t \in [0, 1]$  is the generation probability for timestep  $t$ :

$$\lambda_t = \sigma(w_e^T c_t^x + u_e^T s_t + v_e^T y_{t-1}) \quad (12)$$

The loss function  $\mathcal{L}$  is the average negative log likelihood of the ground-truth target word  $y_t$  for each timestep  $t$ :

$$\mathcal{L} = -\frac{1}{T} \sum_{t=0}^T \log P(y_t) \quad (13)$$

### 2.3 Only-Copy Strategy for Attribute Words

To avoid generating summaries inconsistent with the product attributes, we constrain the attribute words can be presented in the summary only through copying from the source, so that the wrong attribute words cannot be generated. To achieve this goal, for each attribute word  $y_{att}$ , we set  $P_{gen}(y_{att}) = \epsilon$  in Equation 11, where  $\epsilon = 1e-9$ .

We design a heuristic method to collect the attribute words. We find through data analysis that a attribute word in the target is tend to be present in the source. Thus, for each source-target pair, we retrieve each target word in the source, and extract mismatched words as the general word candidates. To guarantee the precision of general word extraction, we regard Chinese words with character-level intersections as the matched words. We regard the target words that are almost never recognized as the general word candidates as the attribute words. The set of attribute words will be released along with our dataset.

	Home Appliances			Bags		
	RG-1	RG-2	RG-L	RG-1	RG-2	RG-L
Lead	30.10	10.43	18.48	25.82	7.78	16.00
Seq2seq	32.51	11.48	<b>21.95</b>	<b>33.36</b>	10.24	21.96
PGNet	32.38	11.47	21.64	33.33	10.62	<b>22.00</b>
PGNet + C-AttriTable	31.47	11.21	21.18	32.67	10.53	21.86
PGNet + AttriTable	32.35	11.59	21.61	33.30	10.51	21.94
PGNet + AttriVocab	32.52	12.15	21.70	33.12	10.62	21.74
PGNet + AttriTable + AttriVocab	<b>32.67</b>	<b>12.21</b>	21.88	33.01	<b>10.94</b>	21.92

Table 1: Experimental results on dataset.

	Faithfulness	Readability
PGNet	64.33%	4.39
PGNet + AttriTable	72.00%	4.43
PGNet + AttriVocab	80.33%	4.37
PGNet + AttriTable + AttriVocab	83.33%	4.40

Table 2: Human Evaluation.

### 3 Dataset

We collect the dataset from a mainstream Chinese e-commerce platform. Each sample is a (*product textual description, product attribute table, product summary*) triplet. The product summaries are generated by thousands of qualified experts, and the auditing groups of the e-commerce platform strictly verify the quality. We collect 361,158 and 92,886 samples for the categories of *Home Appliances* and *Bags*, respectively. For *Home Appliances*, we randomly select 10,000 samples as the validation set and test set. For *Bags* category, we randomly select 5,000 samples as the validation set and test set. The remaining samples are used as the training set. The average number of the Chinese characters for the source text and the summary are 325.54 and 79.24, respectively. The average count of product attribute is 13.87.

## 4 Experiment

### 4.1 Experimental Results

We compare the following baselines. **Lead** is a simple baseline that takes the first 79 characters of the input as the summary. **Seq2seq** model is a standard attention-based seq2seq model without copy mechanism. **PGNet** is the Pointer-Generator Network. **C-AttriTable** denotes concatenate the source text and attribute words in attribute table as the input. **AttriTable** denotes the dual-copy mechanism. **AttriVocab** denotes only-copy strategy for attribute words. Details about experiment settings can be found in our code.

Table 1 shows the results for different models. Generally, the PGNet with attribute tables and the only-copy strategy for attribute words achieves the highest ROUGE score. Although “AttriTable” and “AttriVocab” aim to improve the faithfulness, they exhibit a acceptable performance for the ROUGE score. Considering that ROUGE evaluations are criticized for its poor correlation with human judgment especially for evaluating correctness of the generated summaries (Novikova et al., 2017; Kahn Jr et al., 2009), we perform a human evaluation towards faithfulness and readability.

### 4.2 Human Evaluations

Three expert annotators are involved to evaluate the faithfulness with 0 or 1 and the readability ranging from 1 to 5 (5 is the best) for 100 instances sampled from the test set. From Table 2, we can find that only 64.33% of the summaries are faithful to the source for PGNet, illustrating that faithfulness is an urgent problem for e-commerce product summarization task. “AttriTable” and “AttriVocab” strategies solve the unfaithfulness to a large extend. For the readability, all models obtain comparative results, and we can conclude that our “AttriTable” and “AttriVocab” strategies do not influence the fluency. The statistical significance test using a two-tailed paired t-test gets p-value<0.01.

## 5 Conclusion

We present an e-commerce product summarization model that aims to improve the consistency between the generated summary and the product attributes. We propose two strategies. First, we introduce a dual-copy mechanism that can selectively copy words from both the textual product descriptions and the product attribute table, which makes our model inclined to produce faithful attribute words existing in the product attribute table. Second, we design a heuristic method to recognize the attribute words, and unfaithful attribute words are not allowed to be presented in the summaries through generating from the target vocabulary. We construct a large-scale Chinese e-commerce product summarization dataset, and our dataset and code are available<sup>1</sup>.

## 6 Acknowledgments

This work is partially supported by National Key R&D Program of China (2018YFB2100802).

## References

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *International Conference on Learning Representations*.
- Qibin Chen, Junyang Lin, Yichang Zhang, Hongxia Yang, Jingren Zhou, and Jie Tang. 2019. Towards knowledge-based personalized product description generation in e-commerce. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, pages 3040–3050.
- Sumit Chopra, Michael Auli, and Alexander M. Rush. 2016. Abstractive sentence summarization with attentive recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 93–98, San Diego, California, June. Association for Computational Linguistics.
- Vijay Daultani, Lasguido Nio, and Young-joo Chung. 2019. Unsupervised extractive summarization for product description using coverage maximization with attribute concept. In *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, pages 114–117. IEEE.
- Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O.K. Li. 2016. Incorporating copying mechanism in sequence-to-sequence learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1631–1640, Berlin, Germany, August. Association for Computational Linguistics.
- Charles E Kahn Jr, Curtis P Langlotz, Elizabeth S Burnside, John A Carrino, David S Channin, David M Hovsepian, and Daniel L Rubin. 2009. Toward best practices in radiology reporting. *Radiology*, 252(3):852–856.
- Chandra Khatri, Gyanit Singh, and Nish Parikh. 2018. Abstractive and extractive text summarization using document context vector and recurrent neural networks. *arXiv preprint arXiv:1807.08000*.
- 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 COLING*, pages 1430–1441.
- Haoran Li, Peng Yuan, Song Xu, Youzheng Wu, Xiaodong He, and Bowen Zhou. 2020a. Aspect-aware multi-modal summarization for chinese e-commerce products. In *AAAI*, pages 8188–8195.
- Haoran Li, Junnan Zhu, Jiajun Zhang, Chengqing Zong, and Xiaodong He. 2020b. Keywords-guided abstractive sentence summarization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8196–8203.
- Tianyu Liu, Fuli Luo, Qiaolin Xia, Shuming Ma, Baobao Chang, and Zhifang Sui. 2019. Hierarchical encoder with auxiliary supervision for neural table-to-text generation: Learning better representation for tables. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6786–6793.

---

<sup>1</sup><https://github.com/ypnlp/coling>

- Jekaterina Novikova, Ondřej Dušek, Amanda Cercas Curry, and Verena Rieser. 2017. Why we need new evaluation metrics for NLG. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2241–2252, Copenhagen, Denmark, September. Association for Computational Linguistics.
- 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, September. 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, July. Association for Computational Linguistics.
- Chao Shang, Yun Tang, Jing Huang, Jinbo Bi, Xiaodong He, and Bowen Zhou. 2019. End-to-end structure-aware convolutional networks for knowledge base completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3060–3067.
- Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In *Advances in neural information processing systems*, pages 2692–2700.
- Min Yang, Qiang Qu, Ying Shen, Qiao Liu, Wei Zhao, and Jia Zhu. 2018. Aspect and sentiment aware abstractive review summarization. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1110–1120, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.
- Jiajun Zhang, Yang Zhao, Haoran Li, and Chengqing Zong. 2018. Attention with sparsity regularization for neural machine translation and summarization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(3):507–518.
- 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, July. Association for Computational Linguistics.