



ARTICLE

# Topic-Aware Abstractive Summarization Based on Heterogeneous Graph Attention Networks for Chinese Complaint Reports

Yan Li<sup>1</sup>, Xiaoguang Zhang<sup>1,\*</sup>, Tianyu Gong<sup>1</sup>, Qi Dong<sup>1</sup>, Hailong Zhu<sup>1</sup>, Tianqiang Zhang<sup>1</sup> and Yanji Jiang<sup>2,3</sup>

<sup>1</sup>Engineering Technical Department, China FAW Group Co., Ltd., Changchun, 130012, China

<sup>2</sup>Suzhou Automotive Research Institute, Tsinghua University, Suzhou, 215200, China

<sup>3</sup>College of Software, Liaoning Technical University, Fuxin, 125105, China

\*Corresponding Author: Xiaoguang Zhang. Email: zhangxiaoguang1@faw.com.cn

Received: 20 March 2023 Accepted: 21 July 2023 Published: 08 October 2023

## ABSTRACT

Automatic text summarization (ATS) plays a significant role in Natural Language Processing (NLP). Abstractive summarization produces summaries by identifying and compressing the most important information in a document. However, there are only relatively several comprehensively evaluated abstractive summarization models that work well for specific types of reports due to their unstructured and oral language text characteristics. In particular, Chinese complaint reports, generated by urban complainers and collected by government employees, describe existing resident problems in daily life. Meanwhile, the reflected problems are required to respond speedily. Therefore, automatic summarization tasks for these reports have been developed. However, similar to traditional summarization models, the generated summaries still exist problems of informativeness and conciseness. To address these issues and generate suitably informative and less redundant summaries, a topic-based abstractive summarization method is proposed to obtain global and local features. Additionally, a heterogeneous graph of the original document is constructed using word-level and topic-level features. Experiments and analyses on public review datasets (Yelp and Amazon) and our constructed dataset (Chinese complaint reports) show that the proposed framework effectively improves the performance of the abstractive summarization model for Chinese complaint reports.

## KEYWORDS

Text summarization; topic; Chinese complaint report; heterogeneous graph attention network

## 1 Introduction

The goal of text summarization is to produce a concise sequence of words that captures the most important information from the original document. Most research on text summarization focuses on news articles (short documents), chat logs (multi-party documents), reviews (unstructured documents) and scientific reports (long documents). However, complaint reports are collected daily without receiving much attention. Currently, only 40 government staff are responsible for addressing the reflected problems raised in these reports [1]. Moreover, these reports often use negative language



and oral expressions to describe the cause, process, result, and purpose of events [2]. As shown in Fig. 1, complaint reports contain numerous redundant and meaningless sentences, such as repetition and vague language. Therefore, an automatic summarization model is necessary to reduce redundancy and extract relevant information. This is essential for assisting the government in addressing social problems.

<b>Complaint Report</b> (Translated from Chinese)
The <b>property</b> was scheduled to be delivered in October 2021. The <b>gas</b> needed to be turned on, but the <b>property management</b> and <b>gas company</b> were blaming each other. Finally, the property management replied that the developer's handover inspection had not been completed thoroughly, so the gas cannot be turned on. As a buyer, it was not easy to purchase a property. We received the property normally and paid the <b>property management fees</b> . Now they are telling us that the inspection has not passed and the gas cannot be turned on. Is this reasonable? Why can't the various livelihood issues be resolved before delivery? Now they are blaming each other. As a vulnerable group of people, we can only seek help from the government to give us justice. Name of the community: Hunhe City, Building 6. Is there property management? Has the property management fee been paid?
<b>Summary</b>
A <b>property</b> delivery scheduled for October 2021 has issues with turning on the <b>gas</b> . There is a conflict between the <b>property management</b> and <b>gas company</b> about who's responsible. The buyer is frustrated and seeking government help for justice. The community's name is Hunhe City, Building 6, but it's unclear if there is property management or if the <b>property management fee</b> has been paid.

**Figure 1:** An example of the Chinese complaint report

Existing summarization models for complaint reports fall into two categories: extractive and abstractive. Extractive models aim to select several sentences as a summary, while abstractive models generate a sequence of words that represent the entire document. However, Chinese complaint reports consist of many elliptical sentences that are strongly related to context. Topics can switch frequently and be diverse in each sentence. Therefore, due to the unique characteristic of Chinese complaint reports, extractive models are unavailable to capture all the relevant information in the document whilst the summary maintains a high compression ratio. To tackle this problem, we propose a framework that relies on abstractive summarization with fine-grained topic modeling inspired by review summarization. The main contributions of this paper are summarized as follows:

- To our knowledge, we are the first ones to construct a heterogeneous graph for Chinese complaint reports. A heterogeneous graph with word and topic nodes is constructed to obtain fine-grained textual features. The graph embedding over the heterogeneous graph is proposed to iteratively update the word node.
- A heterogeneous graph attention network with topic distribution is proposed for abstractive summarization.
- We further collect large-scale Chinese complaint reports dataset from the 12345 hotline and platform. Experiments on review and Chinese complaint report datasets show that our proposed model achieves competitive results compared to the state-of-the-art summarization models in different aspects.

## 2 Related Work

### 2.1 Topic Extractor for Text Summarization

Topic signature has been established as a critical element to improve automatic text summarization and information retrieval. However, its implementation in text abstractive summarization is limited. Encapsulating the topic as a representation of the entire document into word embedding is a potential solution for enhancing the sequence network. The word topic distribution of the Latent Dirichlet Allocation (LDA) model can combine with a sequence-to-sequence model to enhance abstractive sentence summarization. Several works reported significant improvements by leveraging topic signatures for text summarization [3–9]. These techniques utilize topic models as an additional mechanism to improve text generation, such as obtaining semantic and sequential features of text generation using topic models [4], mining cross-document subtopics using a topic-based model [5], enriching word representation with topical information [6], and capturing sparse candidate topics using a neighborhood preserving semantic measure [7]. Ailem et al. [8] developed a topic-augmented decoder that generates a summary conditioned on both the input document and the latent topics of the document. They found that latent topics reveal more global semantic information that can be used to bias the decoder to generate words. Gao et al. [9] incorporated a neural generative topic matrix as an abstractive level of topic information by mapping global semantics into a local generative language model. Inspired by the success of topic modeling, we propose a summarization framework that uses latent topics to capture long-range dependencies in documents.

### 2.2 Text Summarization Based on Graph

Graph neural networks (GNN) have become increasingly popular for modeling relationships between text spans for abstractive summarization. Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT) are two representative GNN models that aggregate node embeddings to obtain compressed graph representations. While GNNs are originally designed to represent the entire document with the same node type, multi-type nodes in the graph have emerged as novel applications for text summarization. Recent work in text summarization has explored how to model these structures. Wang et al. [10] constructed a heterogeneous graph using GAT to learn cross-sentence relations and converted it to multi-document extractive summarization. Cui et al. [11] integrated a joint Neural Topic Model (NTM) to discover latent topics, which can provide document-level features for sentence selection. Jia et al. [12] combined heterogeneous and homogeneous graphs to represent multi-level features. Li et al. [13] proposed an extension of Graph Neural Networks to incorporate graph-structured information into output sequences, achieving remarkable results on bAbI tasks [14]. Ma et al. [15] proposed a graph-based semi-supervised learning model for document classification. Similarly, Yao et al. [16] used GCN for text classification, constructing a large graph containing both words and documents. Inspired by previous work on graph-based models, we propose a heterogeneous graph to represent Chinese complaint reports. Moreover, the representation of topic and word embedding is updated by message passing in the heterogeneous graph attention network.

## 3 Method

To generate a summary with less redundancy and suitable information, the proposed model consists of 4 components: topic extractor, word embedding, graph embedding, and sentence decoder, as shown in Fig. 2. Firstly, the topic extractor aims to extract the topic distribution of the entire document and each word. Secondly, similar to other summarization models, word embedding is to learn the contextual representation of each word. Additionally, graph embedding uses a heterogeneous

graph to combine global features (topic distribution) and local features (contextual representation). Finally, the decoder generates a summary based on the contextual representation. Further details of each architecture are provided in the following section.

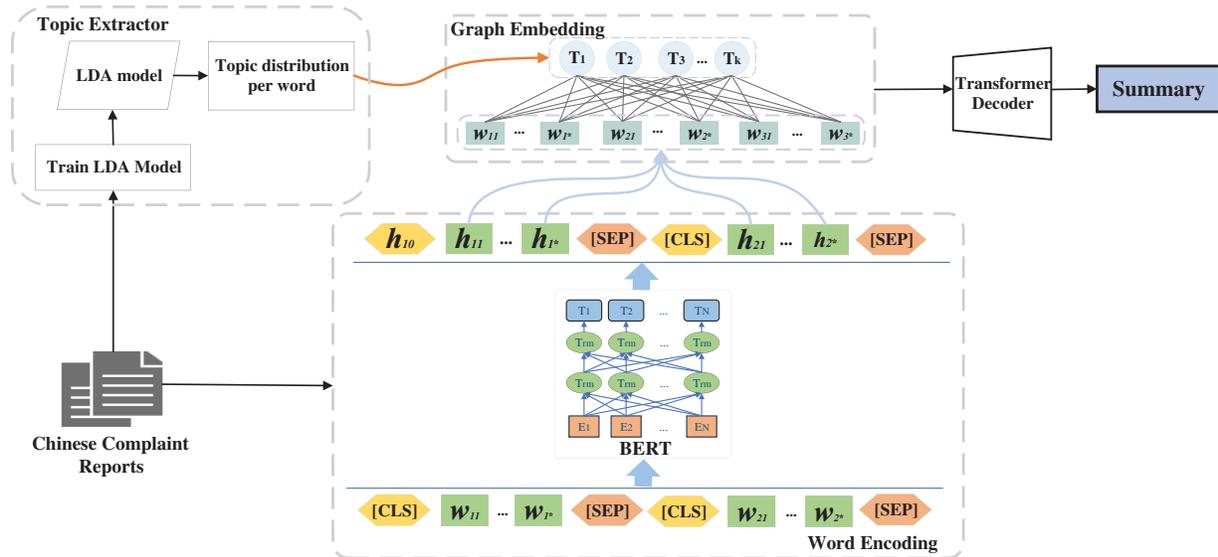


Figure 2: Schema of our proposed model architecture

### 3.1 Problem Definition

Formally, a Chinese complaint report contains  $n$  sentences  $D = \{s_1, s_2, \dots, s_n\}$ , where the  $i$ -th sentence can be represented as  $s_i = \{w_{i1}, w_{i2}, \dots, w_{im}\}$ . The abstractive summarization process for complaint reports is formulated as a text-generation task, which aims at generating an abstractive summary  $S = \{w_1, w_2, \dots, w_r\}$  using a heterogeneous graph. In detail, the sequence-to-graph transformation learns to construct a graph structure that captures the global and local features of the original document. The graph-based encoder generates contextual hidden states with graph features denoted as  $H = \{H_1, H_2, \dots, H_n\}$ . The decoder combines updated word embeddings to form a summary  $S = F(H) = \{w_1, w_2, \dots, w_r\}$ , where  $F(\cdot)$  is an abstractor function. Meanwhile, the gold-standard summary contains  $l$  words  $S_G = \{a_1, a_2, \dots, a_l\}$ .

### 3.2 Topic Extractor

Topic, as a high-level semantic unit, becomes a latent feature of the whole document. In this paper, a topic extractor model based on a probabilistic distribution is employed to enrich the semantic representation of Chinese complaint reports. Each topic can be seen as a distribution of semantically coherent terms and each document exhibits these topics with different probabilities or proportions. The main purpose of the topic extractor model is to transform the whole document into latent topic features. Compared to other topic approaches (such as Latent Semantic Analysis, Probability Latent Semantic Analysis), LDA [17] can generate a well-topic distribution and reduce the risk of model overfitting. Therefore, LDA as the topic model to generate a topic vector  $T_{w_j}$  for each word  $w_j$  is chosen. Meanwhile, the topic distribution  $T_D$  of the whole document is obtained in the same way. In this paper, we consider that the topic representation of the whole document conveys the topic of words. The topic

vector of the word  $w_j$  can be represented as:

$$T_j = T_{w_j} + T_D \quad (1)$$

### 3.3 Word Embedding

Word embedding is computed for each word to capture contextual information. A pre-training model is employed to represent the local features of each word as a structural content vector. The original document is first represented as word embeddings using the most advanced pre-trained language model, Bidirectional Encoder Representations from Transformers (BERT), which greatly guarantees the contextual semantics. BERT trains on a corpus of 33 million words via masked language modeling and next-sentence prediction. Formally, the input text is first preprocessed by inserting two special tokens. The token ‘[CLS]’ is inserted at the beginning of each sentence, and the output calculated by this token is used to integrate the information in each sequence. Meanwhile, the token ‘[SEP]’ is inserted at the end of each sentence as an indicator of sentence boundaries. Specifically, the special tokens ‘[CLS]’ and ‘[SEP]’ are inserted at the beginning and end of each sentence, respectively. The hidden states can be represented as follows:

$$\{h_{10}, h_{11}, \dots, h_{1*}, \dots, h_{n0}, h_{n1}, \dots, h_{n*}\} = BERT(w_{10}, w_{11}, \dots, w_{1*}, \dots, w_{n0}, w_{n1}, \dots, w_{n*}) \quad (2)$$

where  $w_{ij}$  is the  $j$ -th word in the  $i$ -th sentence.

### 3.4 Graph Embedding

**Heterogeneous graph construction** Given a complaint report  $D$ , we start with modeling  $D$  as a semantic graph. The constructed graph can be defined as  $G = \{V, E\}$ . Topic words  $T_D$  and tokens (words in  $D$ )  $w_j$  can be considered graph nodes  $V$ . Meanwhile, the two edge types are created: 1) inner-connection edge  $e_{w-w}$ , which represents the word dependency in a sentence. The edge weight for a word pair is the cosine similarity of their embedding. We choose GloVe as the static word embedding in this paper. 2) topic-word edge  $e_{t-w}$ , which connects nodes of topics and semantic words. The edge weights of topic-word are set to 1.0 to preserve the hierarchical structure. Therefore, the constructed heterogeneous graph can simultaneously capture the global and local features since the topic and word can represent the substance of the whole document and the dependency relationship, respectively.

**Graph Encoder** Node embedding for the heterogeneous graph is initialized using contextual embedding and topic distribution. As the previous graph neural network cannot deal with the heterogeneous graph through message passing, a heterogeneous graph attention network is proposed, inspired by the graph attention network. The network aggregate embeddings of 2 edge types (word-word and word-topic edge). The hidden state of a word  $W_i$  in the  $(l+1)$ -th layer is shown as follows:

$$h_i^{l+1} = UH_i^l \quad (3)$$

$$H_i^l = h_i^{l:w-w} \cdot h_i^{l:w-e} \quad (4)$$

where  $h_i^{l:w-w}$  and  $h_i^{l:w-e}$  are the updated word embedding of word-word and word-topic edge type, respectively.  $H_i^l$  represents the output embedding of node  $i$  at the  $l$ -th layer.

**Graph Propagation** We initial the representation of each node by combining topic extractor  $H_T \in \mathbb{R}^{m \times d_T}$  and word embedding  $H_W \in \mathbb{R}^{n \times d_W}$ , as follows:

$$H_i^{(0)} = h_i^W \parallel h_i^T \quad (5)$$

where  $d_W$  and  $d_T$  are the dimension of word embedding and topic representation, respectively.

Then, the representation of semantic nodes in the proposed heterogeneous graph can be updated with the graph attention network. The updating process of the graph attention layer with multi-head attention is designed as follows:

$$Z_{i,j} = \text{LeakyReLU}(W_a[W_q h_i; W_k h_j]) \quad (6)$$

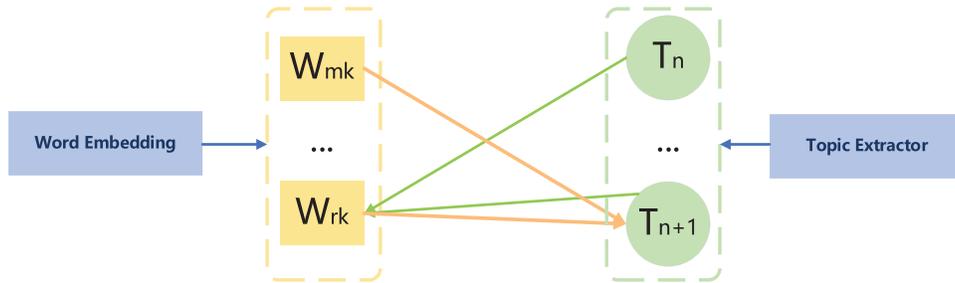
$$\alpha_{i,j} = \frac{\exp(Z_{i,j})}{\sum_{l \in N_i} \exp(Z_{i,l})} \quad (7)$$

$$\mu_i = \sum_{j \in N_i} \sigma(\alpha_{i,j} W_c h_j) \quad (8)$$

$$h'_i = \parallel_{k=1}^K \sum_{j \in N_i} \sigma(\alpha_{i,j}^k W_c^k h_j) \quad (9)$$

where  $W_a, W_q, W_k$  and  $W_c$  are the model trainable parameters.  $K$  is the number of attention heads.  $N_i$  represents the neighbors of  $i$ -th node.  $\alpha_{i,j}$  is the attention weight between  $h_i$  and  $h_j$ .

Topic and word vectors are updated through the proposed graph propagation, shown in Fig. 3. On the one hand, topic vectors, as the intermediary to build an inter-word relationship, can be considered as the global feature. On the other hand but similarly, updated word embedding as the local feature not only captures the contextual feature but also distills the high-level semantic units from topic vectors. The graph attention layer connects the semantic relationship between topics and words to pay attention to the salient features.



**Figure 3:** The processing of graph propagation

Moreover, the graph attention layer is further modified to infuse the edge information weights ( $e_{w-w}$  and  $e_{t-w}$ ), which can be represented as a matrix in multi-dimensional embedding space ( $e_{w-w} \in \mathbb{R}^{m \times d_w}$  and  $e_{t-w} \in \mathbb{R}^{m \times d_T}$ ). Therefore, Eq. (3) is updated as follows:

$$Z_{i,j} = \text{LeakyReLU}(W_a[W_q h_i; W_k h_j; e_{w-t}; e_{w-w}]) \quad (10)$$

### 3.5 Decoder

The decoder of the proposed summarization model employs a Transformer architecture and uses topic-aware word representation and contextual information to predict the probability of each word in the summary sequence. During decoding, the beam search algorithm is utilized to generate a group tag at each search step to construct the final summary. In this way, the model can better capture the theme and relevant information of the original text, thus generating a more accurate and comprehensive text summary  $Y = \{Y_1, Y_2, \dots, Y_r\}$ .

### 4 Experiments and Analysis

In this section, we first present the Chinese complaint reports and public review (Yelp and Amazon) datasets and hyperparameters setting. Then experimental results on these datasets are introduced to verify the performance of our proposed model. Lastly, the ablation and case studies are given for further analysis.

#### 4.1 Dataset

The Chinese complaint reports are collected from the government service hotline and website where complaints have taken place between the government department and complainer in the Chinese language. The Chinese complaint reports dataset contains 220,855 complaint reports that happened in recent 2 years. For the raw data, we run a preprocessing to replace the invalid data (such as consulting, or dialing a wrong number). Additionally, all the complaint-summary pairs are manually annotated with 3 government department staff. To demonstrate the complexity of Chinese complaint reports, the problem types are shown in Fig. 4. Moreover, we select 100–500 Chinese characters as the proper length of Chinese complaint reports. The length distribution of the Chinese complaint report is demonstrated in Fig. 5. The results are reported in F1 ROUGE scores, containing ROUGE-1 (unigram), ROUGE-2 (bigram) and ROUGE-L (Longest Common Subsequence, LCS).

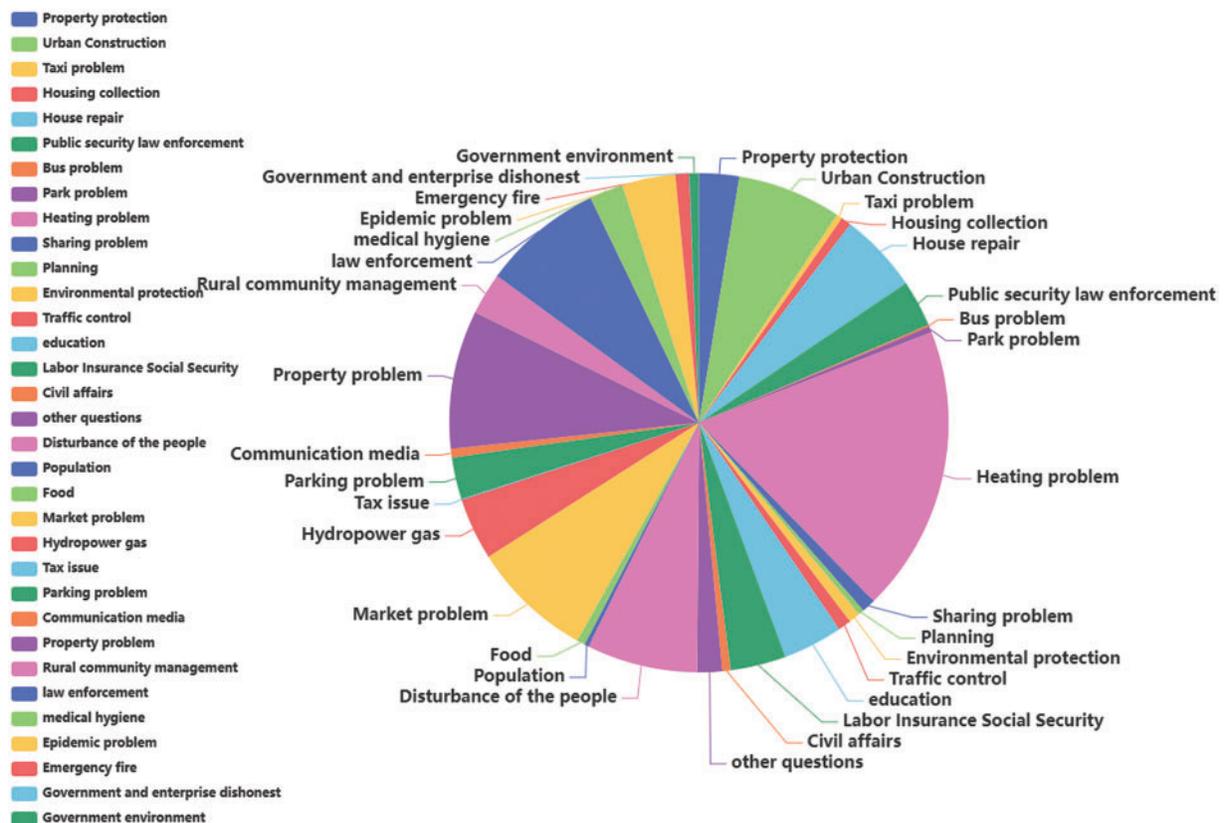
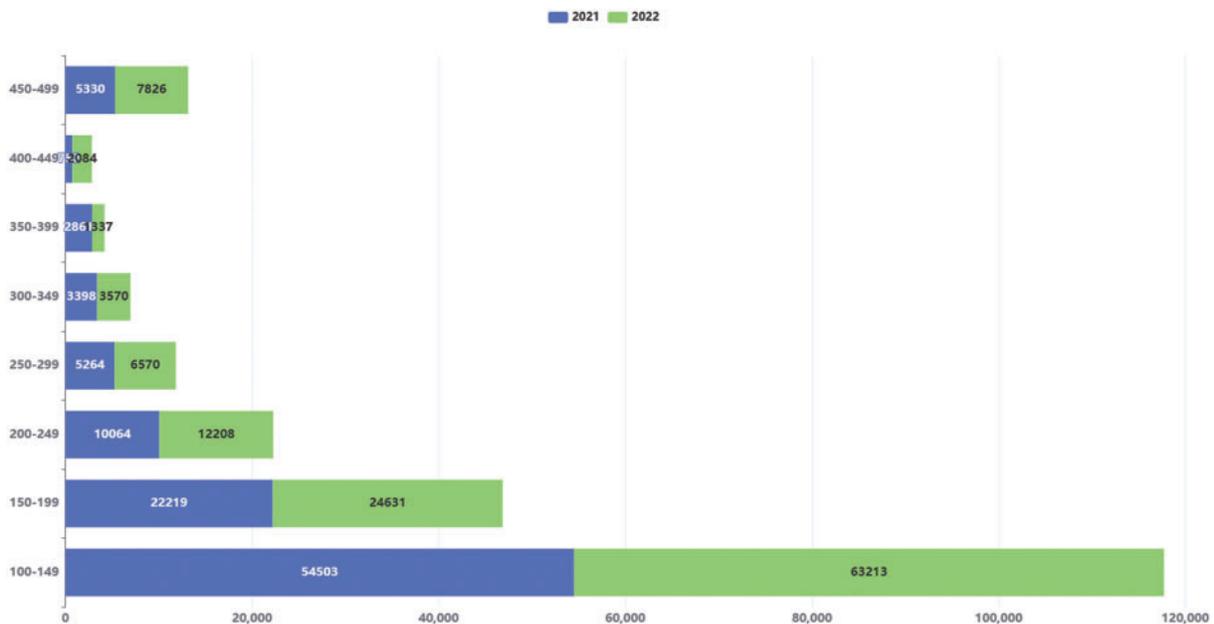


Figure 4: The problem type of Chinese complaint reports



**Figure 5:** The length distribution of Chinese complaint reports

#### 4.2 Hyperparameter Setting

The dimension of word embedding and topic vector is initialized as 768. The transformer decoder is set to 2 layers. Specifically, Transformer has 768 hidden units and 8 heads. We utilize the Adam optimizer [18] to optimize the parameters in the proposed model. The learning rate is 0.001. The word embedding is set to the pre-trained Chinese BERT model released by [19]. We train our proposed model for 5,000 epochs. Except for the pre-trained language model, the other parameters are randomly initialized.

#### 4.3 Comparison Methods

Several state-of-the-art comparison models are applied to evaluate our proposed model:

- Lead [20]: Lead is to select the first several sentences in a document as the summary.
- Oracle [20]: It uses the greedy algorithm to select the most salient sentences as the gold summary.
- TextRank [21]: It transforms the original document into a graph and selects several sentences through a graph-based ranking algorithm.
- PacSum [22]: It improves the performance of TextRank with the edge-weighted calculation in a graph.
- MeanSum [23]: Through an autoencoder model, it calculates the representation in average to decode a summary.
- BERTSUM [24]: It inserts multiple segmentation tokens into the original document to obtain each sentence representation. It is the first BERT-based extractive summarization model.
- HiBERT [25]: It modifies BERT into a hierarchical structure and designs an unsupervised method to pre-train it.

- DISCOBERT [26]: It is a state-of-the-art BERT-based extractive model which encodes documents with BERT and then updates sentence representations with a graph encoder. DISCOBERT builds a document graph with only sentence units based on discourse analysis.
- ConsistSum [27]: An unsupervised opinion summarization system can capture the consistency of aspects and sentiment between reviews and summaries. It first constructs high-quality synthetic datasets and then utilizes them to train the summarization model.
- TranSum [28]: TranSum is an abstractive summarization model that learns embeddings of reviews by utilizing intra-group and inter-group invariances.

#### 4.4 Automatic Evaluation

The compared results on the Chinese complaint report dataset are shown in Table 1. The comparison methods can be categorized into three groups: baselines, extractive models and abstractive models.

**Table 1:** The compared results on the Chinese complaint report dataset

Type	Methods	ROUGE-1	ROUGE-2	ROUGE-L
Baselines	Lead	27.12	10.32	21.60
	Oracle	52.48	22.98	35.91
Extractive summarization	TextRank	28.27	11.50	22.63
	PacSum	29.18	12.04	23.07
	BERTSUM	33.59	13.57	26.25
	HiBERT	36.10	14.81	27.58
	DISCOBERT	37.49	15.75	27.04
Abstractive summarization	MeanSum	36.83	16.62	28.61
	ConsistSum [27]	37.32	15.94	28.41
	TranSum [28]	38.23	16.24	27.49
	Proposed model	<b>40.84</b>	<b>17.27</b>	<b>29.52</b>

Table 1 depicts Oracle and Lead baselines in the first part, followed by state-of-the-art extractive summarization models in the second part. In the third part, abstractive methods are presented, including DISCOBERT which employs a topic model and pre-training language model BERT. Notably, our proposed model achieves highly competitive results. Compared to extractive models, our proposed graph-based model with topic features demonstrates superior performance. The result proves that the proposed graph structure with topic features is positively effective for the summarization model performance. While most extractive models select several sentences containing salient information, our proposed model with a heterogeneous graph outperforms other extractive models, ensuring both informativeness and conciseness of the generated summary. Moreover, our proposed model achieves competitive performance against other abstractive models on all metrics, affirming the effectiveness of high-level topic features and heterogeneous graph construction.

Furthermore, to understand how much  $k$  affects the generated summary, we compare the topic number setting of the topic extractor in Table 2. Considering the characteristic of Chinese complaint reports, we set the topic number to 2, 4, 6, 8, 10. The result demonstrates that the quality of generated

summary decreases with the increasing topic number. We infer that the topic feature is more unobvious with the more topic number. Moreover, the result manifests that the topic influences the performance of the summary model. Therefore, we set the topic number to 4.

**Table 2:** The influence of different topic numbers on complaint summarization

Topic number	ROUGE-1	ROUGE-2	ROUGE-L
2	37.62	14.16	25.98
4	<b>40.84</b>	<b>17.27</b>	<b>29.52</b>
6	39.81	17.06	28.56
8	38.1	16.32	27.43
10	36.81	14.58	26.82

To verify the generalization of our proposed framework, Yelp and Amazon datasets are utilized to obtain the review summary. The result of 2 review datasets is depicted in Table 3. For the Yelp dataset, our framework keeps a similar performance of SOTA abstractive summarization (ConsistSum). For the Amazon dataset, our proposed model outperforms all of the compared benchmark summarization models, demonstrating the topic feature and heterogeneous graph embedding improve the result of generated summary. This shows that our model is also suitable for the review datasets.

**Table 3:** Model performance on Yelp and Amazon datasets

Models	Yelp			Amazon		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
Opinosis	25.15	2.61	13.54	28.42	4.57	15.50
MeanSum	28.86	3.66	15.91	29.20	4.70	18.15
DISCOBERT	29.30	5.77	18.56	–	–	–
PlanSum	30.14	4.99	17.65	32.87	6.12	19.05
TranSum	–	–	–	34.23	7.24	20.49
ConsistSum	<b>32.65</b>	<b>7.49</b>	<b>20.87</b>	33.32	5.94	21.41
Our framework	31.59	7.08	19.94	<b>34.85</b>	<b>7.64</b>	<b>21.86</b>

#### 4.5 Human Evaluation

Intuitively, our proposed model can paraphrase abstractive summarization instead of extracting several sentences, improving the conciseness of the summary. Furthermore, the proposed abstractive summarization model with high-level semantic features can enhance readability and informativeness. To evaluate the hypothesis, human evaluation is conducted by 3 government experts. The 100 random samples from the test set are selected. The evaluation scores contain conciseness, readability and informativeness of summaries, which are calculated from 1 (the worst) to 5 (the best) on average.

The result of the human evaluation is shown in Table 4. We compare our model against TextRank, BERTSum, HiBERT and DISCOBERT. The proposed model produces high-quality summaries accepted by volunteers. It means that the summary generated by the proposed model is more relevant

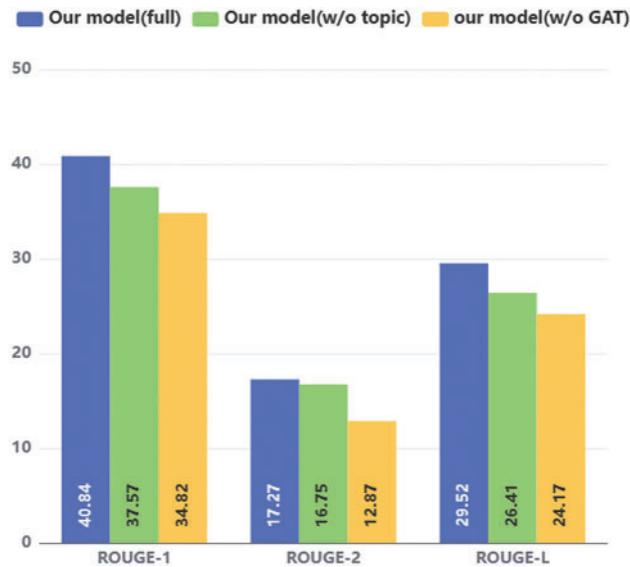
than others to the original report. Meanwhile, the conciseness result of generated summary signifies that the proposed abstractive model obtains the suitable length through the transformer decoder. It indicates that our model can generate the most relevant summary covering different topics and effectively reduce the redundancy of generated summary. Finally, in terms of readability, our model weakly outperforms other models since the other models are extractive models which select several sentences from the original document to keep the readability.

**Table 4:** The human evaluation result of comparison models for complaint summarization

Methods	Informativeness	Conciseness	Readability
TextRank	3.42	3.04	3.71
BERTSUM	4.17	3.24	3.76
HiBERT	4.28	3.39	3.80
DISCOBERT	4.30	3.17	3.85
Proposed model	<b>4.51</b>	<b>3.68</b>	<b>3.92</b>

#### 4.6 Ablation Study

We also perform an ablation study to evaluate each module in the proposed model. Fig. 6 shows the results of each part in the proposed framework. To show the effectiveness of topic features and graph attention layer, we calculate the F1 ROUGE scores of the proposed model without topic features and the heterogeneous graph.



**Figure 6:** The ablation result of our proposed model

The model without topic features means that the homogeneous graph is based on words. The model without GAT represents that the representation of word embedding is the combination of contextual and topic information. Through the ablation study, the proposed model without topic features obtains more salient performance than the model without GAT. The result demonstrates that

graph propagation can combine the contextual information into word embedding which is effective for the abstractive summary model. Therefore, it can be seen that the GAT in the proposed model is more capable of summary generation.

#### 4.7 Case Study

Table 5 provides an example to illustrate the role of topic features in enhancing our proposed summarization model. The table includes the original complaint report, its gold summary, the best extractive summary, the summary generated by SOTA models and the summary generated by our proposed model. The report is divided into four topics—community, heating, temperature and a problem highlighted in blue. While the extractive model summary contains informative content, it is longer than both the gold summary and the summary generated by our proposed model. Additionally, our proposed model generates a summary that captures important information related to the topic. Compared to state-of-the-art models such as MeanSum and ConsistSum, our proposed model generates the shortest summary, indicating that it achieves the right balance between conciseness and informativeness. This suggests that the topic-aware approach employed in our proposed model effectively preserves salient information while keeping the summary length.

**Table 5:** A case study of the proposed model

Complaint report	I am a resident of the 4th floor, Unit 1, Building 20, in Tongcheng Yipin Community. The community was built in 2010. I moved into the community in 2019. The heating unit of the community is Heyi Water Heating Co., Ltd., Fusun Economic Development Zone (China). Since the beginning of the heating, the temperature in my home is lower than 18 degrees. The city property community heating is not in place causing district heating has not been good. Many residents of the Community group reflect that the elderly cannot live due to the low temperature. Please urge the government departments of the demonstration area to supervise the heating units. Solve the problem of substandard heating as soon as possible. Refund and compensate for the substandard situation according to the regulations!
Gold summary	Heating substandard requirements as soon as possible to solve and refund in Tongcheng Yipin Community 20 building 1 unit 4 floor.
Extractive model	I am a resident of the 4th floor, Unit 1, Building 20, in Tongcheng Yipin Community. Since the beginning of the heating, the heating temperature in my home is lower than 18 degrees. Please urge the government departments of the demonstration area to supervise the heating units. Solve the problem of substandard heating as soon as possible.
MeanSum	Residents in Tongcheng Yipin Community are facing inadequate heating in their homes due to a poor heating system provided by Heyi Water Heating Co., Ltd., which has caused temperatures to remain below 18 degrees Celsius since moving into the community in 2019. Several elderly residents are unable to bear the low temperatures. The author is urging the government departments of the demonstration area to supervise heating units as soon as possible to solve the heating issue and provide compensation for the substandard situation based on regulations.

(Continued)

**Table 5 (continued)**

ConsistSum	Residents of Tongcheng Yipin Community are facing substandard heating due to a poor heating system provided by Heyi Water Heating Co., Ltd., resulting in temperatures being constantly below 18 degrees Celsius since 2019. Elderly residents in particular have had difficulty living in these conditions. The government departments of the demonstration area are being urged to supervise heating units to quickly solve the issue and provide compensation based on regulations.
Proposed model	In Tongcheng Yipin <b>Community</b> , the <b>heating</b> unit is Heyi Water Heating Co., Ltd. The <b>temperature</b> is low. Urge the government. Please solve the <b>problem</b> . Refund the heating fee.

## 5 Conclusion

In this paper, a novel approach for generating topic-aware abstractive summaries of Chinese complaint reports is presented through a proposed heterogeneous graph attention network. To capture the high-level semantic information that is essential for generating an accurate summary, a topic-aware heterogeneous graph is constructed. Additionally, in terms of the unique characteristics of Chinese complaint reports, topic features (the main content of the original document) as extra nodes in the heterogeneous graph are designated. Our experimental results demonstrate the effectiveness of our model for generating informative and non-redundant summaries of Chinese complaint reports and public review datasets.

**Acknowledgement:** We thank the anonymous reviewers for their useful comments.

**Funding Statement:** This work is partially supported by National Natural Science Foundation of China (52274205) and Project of Education Department of Liaoning Province (LJKZ0338).

**Author Contributions:** The authors confirm contribution to the paper as follows: study conception and design: Yan Li, Xiaoguang Zhang; data collection: Tianyu Gong, Qi Dong, Hailong Zhu; analysis and interpretation of results: Yan Li, Xiaoguang Zhang, Tianyu Gong; draft manuscript preparation: Hailong Zhu, Tianqiang Zhang and Yanji Jiang. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** Reader can email the corresponding authors in order to obtain the available data of Chinese complaint reports.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] X. Peng, Y. Li, Y. Si, L. Xu, X. Liu *et al.*, "A social sensing approach for everyday urban problem-handling with the 12345-complaint hotline data," *Computers, Environment and Urban Systems*, vol. 94, pp. 101790, 2022.

- [2] R. Tao and C. Shuang, “Weakly supervised abstractive summarization with enhancing factual consistency for Chinese complaint reports,” *Computers, Materials & Continua*, vol. 75, no. 3, pp. 6201–6217, 2023.
- [3] H. Pan, H. Liu and Y. Tang, “A Sequence-to-sequence text summarization model with topic based attention mechanism,” in *Proc. of WISA*, Qingdao, China, pp. 285–297, 2019.
- [4] H. Tang, M. Li and B. Jin, “A topic augmented text generation model: Joint learning of semantics and structural features,” in *Proc. of EMNLP-IJCNLP*, Hong Kong, China, pp. 5090–5099, 2019.
- [5] X. Zheng, A. Sun, J. Li and M. Karthik, “Subtopic-driven multi-document summarization,” in *Proc. of EMNLP-IJCNLP*, Hong Kong, China, pp. 3153–3162, 2019.
- [6] S. Narayan, S. B. Cohen and M. Lapata, “Don’t give me the details, just the summary! Topic-aware convolutional neural networks for extreme summarization,” in *Proc. of EMNLP*, Brussels, Belgium, pp. 1797–1807, 2018.
- [7] Z. Yang, Y. Yao and S. Tu, “Exploiting sparse topics mining for temporal event summarization,” in *Proc. of ICIVC*, pp. 322–331, 2020.
- [8] M. Ailem, B. Zhang and F. Sha, “Topic augmented generator for abstractive summarization,” arXiv preprint, arXiv:1908.07026, 2019.
- [9] Y. Gao, Y. Wang, L. Liu, Y. Guo and H. Huang, “Neural abstractive summarization fusing by global generative topics,” *Neural Computing and Applications*, vol. 32, pp. 5049–5058, 2020.
- [10] D. Wang, P. Liu, Y. Zheng, X. Qiu, X. Huang *et al.*, “Heterogeneous graph neural networks for extractive document summarization,” in *Proc. ACL*, pp. 6209–6219, 2020.
- [11] P. Cui, L. Hu and Y. Liu, “Enhancing extractive text summarization with topic-aware graph neural networks,” in *Proc. of the 28th Int. Conf. on Computational Linguistics*, Barcelona, Spain, pp. 5360–5371, 2020.
- [12] R. Jia, Y. Cao, H. Tang, F. Fang, C. Cao *et al.*, “Neural extractive summarization with hierarchical attentive heterogeneous graph network,” in *Proc. of EMNLP*, pp. 3622–3631, 2020.
- [13] Y. Li, D. Tarlow, M. Brockschmidt and R. Zemel, “Gated graph sequence neural networks,” in *Proc. of ICLR*, Caribe Hilton, San Juan, Puerto Rico, pp. 3104–3112, 2016.
- [14] J. Weston, A. Bordes, S. Chopra, A. M. Rush, B. van Merriënboer *et al.*, “Towards ai-complete question answering: A set of prerequisite toy tasks,” arXiv preprint, arXiv:1502.05698, 2015.
- [15] T. Ma, H. Wang, L. Zhang, Y. Tian and N. Al-Nabhan, “Graph classification based on structural features of significant nodes and spatial convolutional neural networks,” *Neurocomputing*, vol. 423, pp. 639–650, 2021.
- [16] L. Yao, C. Mao and Y. Luo, “Graph convolutional networks for text classification,” in *Proc. of AAAI*, Honolulu, Hawaii, USA, pp. 396–404, 2019.
- [17] D. M. Blei, “Probabilistic topic models,” *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012.
- [18] D. P. Kingma and J. B. Adam, “A method for stochastic optimization,” in *Proc. of ICLR*, San Diego, CA, USA, pp. 1–15, 2015.
- [19] Y. Cui, W. Che, T. Liu, B. Qin and Z. Yang, “Pre-training with whole word masking for Chinese bert,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3504–3514, 2021.
- [20] R. Nallapati, F. Zhai and B. Zhou, “Summarunner: A recurrent neural network based sequence model for extractive summarization of documents,” in *Proc. of AAAI*, New York, USA, vol. 31, no. 1, pp. 3075–3081, 2017.
- [21] R. Mihalcea and P. Tarau, “Textrank: Bringing order into text,” in *Proc. of EMNLP*, Barcelona, Spain, pp. 404–411, 2004.
- [22] H. Zheng and M. Lapata, “Sentence centrality revisited for unsupervised summarization,” in *Proc. of ACL*, Florence, Italy, pp. 6236–6247, 2019.
- [23] E. Chu and P. Liu, “MeanSum: A neural model for unsupervised multi-document abstractive summarization,” in *Proc. of ICML*, Long Beach, CA, USA, pp. 1223–1232, 2019.
- [24] L. Yang and L. Mirella, “Text summarization with pre-trained encoders,” in *Proc. of EMNLP-IJCNLP*, Hong Kong, China, pp. 3728–3738, 2019.

- [25] X. Zhang, F. Wei and M. Zhou. “HIBERT: Document level pre-training of hierarchical bidirectional transformers for document summarization,” in *Proc. of ACL*, Florence, Italy, pp. 5059–5069, 2019.
- [26] J. Xu, Z. Gan, Y. Cheng and J. Liu, “Discourse-aware neural extractive text summarization,” in *Proc. of ACL*, Tokyo, Japan, pp. 5021–5031, 2020.
- [27] W. Ke, J. Gao, H. Shen and X. Cheng, “ConsistSum: Unsupervised opinion summarization with the consistency of aspect, sentiment and semantic,” in *Proc. of WSDM*, NY, US, pp. 467–475, 2022.
- [28] K. Wang and X. Wan, “TransSum: Translating aspect and sentiment embeddings for self-supervised opinion summarization,” in *Proc. of ACL*, Punta Cana, Dominican Republic, pp. 729–742, 2021.