Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.036178
Article

Weakly Supervised Abstractive Summarization with Enhancing Factual Consistency for Chinese Complaint Reports

Ren Tao and Chen Shuang*

Software College, Northeastern University, Shenyang, 110000, China
*Corresponding Author: Chen Shuang. Email: chenshuang888@163.com
Received: 20 September 2022; Accepted: 29 December 2022

Abstract: A large variety of complaint reports reflect subjective information expressed by citizens. A key challenge of text summarization for complaint reports is to ensure the factual consistency of generated summary. Therefore, in this paper, a simple and weakly supervised framework considering factual consistency is proposed to generate a summary of city-based complaint reports without pre-labeled sentences/words. Furthermore, it considers the importance of entity in complaint reports to ensure factual consistency of summary. Experimental results on the customer review datasets (Yelp and Amazon) and complaint report dataset (complaint reports of Shenyang in China) show that the proposed framework outperforms state-of-the-art approaches in ROUGE scores and human evaluation. It unveils the effectiveness of our approach to helping in dealing with complaint reports.

Keywords: Automatic summarization; abstractive summarization; weakly supervised training; entity recognition

1 Introduction

Complaint reports of urban management problems, like the 311 hotlines in US cities, have been established by specific government branch [1]. Complaint reports range from large-scale trouble, containing air pollution, vital infrastructure, etc. In order to guarantee the welfare of city residents, it is an essential task for urban managers to effectively and timely solve the problems in complaint reports [2]. Moreover, nearly 400 employees need to record and dispose of 10,000 daily hotline complaints in a city. Needless to say, one would not expect the accuracy and consistency of disposal results with this method, as the results solely rely on the experiences of uncertain 400 employees. Therefore, automatic text summarization helps establish a frontier to address the problem which is numerous unreliable complaint data labels. These complaint reports should be generated in a shorter version and downsized for easier understanding. As a result, it is desirable to propose an automatic summarization model. It can create a summary of complaint reports published by the government department.

Complaint reports are distinct from reviews or opinions. Firstly, a review or opinion (Twitter, Amazon and Weibo) describes a product or event. On the contrary, complaint reports reflect real-time and spatially explicit urban problems which are fine-grained. Secondly, the sentimental polarity of a review is positive, neutral or negative while that of complaint reports almost is negative. Meanwhile, a customer may seek reviews before purchasing a product or gaining an attitude concerning an event. A citizen writes a complaint report to the related website or hotline, and aim to timely solve the existing problem. Considering the difference between the review and complaint reports, it is confusing to easily evaluate and mine the deep information in complaint reports. The purpose of complaint reports is to define an automatic tool that can extract salient information in favor of solving existing problems. Therefore, automatic text summarization can significantly help in benefiting from the semantic information to obtain the salient content. In essence, after producing a summary of complaint reports, the related department can timely and effectively deal with the existing problem. To our knowledge, no study has focused on summarization of complaint reports.

In general, complaint summarization can be extractive, i.e., obtained by identifying the top-k salient sentences from the original text, or abstractive, generated by complaint scratch. However, the selected sentences for extractive summarization contain irrelevant and redundant information. Additionally, the coherence of extracted sentences is weak for extractive summarization models. Meanwhile, an abstractive summary of opinions, reviews, etc. gives a better result than the extractive summarization model. Therefore, an abstractive model is utilized for complaint reports.

The state-of-the-art abstractive summarization models contain lexical, semantic, and syntactic levels along with discourse processing. The selected features [3], rating those features [4] and sentence identification that contains features [5] should be included in the task of complaint reports summarization. Previous researches [6–8] on abstractive summarization show that sequence-to-sequence model, also known as the encoder-decoder model, makes the performance of summarization model continuously improved. After Rush et al. [9] apply the encoder-decoder framework to abstractive summarization, they propose an abstractive summarization with a basic attention mechanism. The most significant advantage of the model is the utilization of powerful attention mechanism and beam search strategy [10]. Following these researches, the hierarchical encoder-decoder models [11–13] have shown strong performance in document language tasks. Although the language models of abstractive summarization have obtained success in model performance, abstractive summarization model still exhibits two limitations:

1. **Factual consistency problem.** Even with the existing abstractive models enabling a better summary result for each complaint text, the further problem of factual consistency is encountered in generated summary.
2. **Label absence problem.** Similar to review summarization, complaint data lacks gold-standard summaries. To address these issues, a novel framework for complaint report summarization task is proposed. Specifically, the structure of complaint report is defined in this paper. Based on the summary structure, the combination of entity and intention is proposed as the label of complaint report. Moreover, fact corrector is utilized in order to keep the factual consistency of complaint reports. Our contributions can be summarized as follows:

1. **Entity-based abstractive model.** Different from previous literature focused on generated summaries, an entity-based abstractive model is proposed which ensures factual consistency in the generated summaries.
2. **A weakly supervised model.** For the problem of label absence and unreliable data, a weakly supervised model with combination strategy is proposed to reduce human-labeled costs for complaint report summarization.
3. **Generalization of the proposed model.** In order to demonstrate the generalization of the proposed model, we evaluate the proposed model on universal review datasets (Yelp and Amazon) and the real complaint data in Shenyang, one of the biggest cities in China.

The rest of this paper is organized as follows. In Section 2, related work is firstly discussed concerning abstractive summarization. We then demonstrate our framework in Section 3. The experiments and analysis are presented in Section 4. Finally, a conclusion is given in Section 5.

2 Related Work

Automatic summarization has been an active area of research. Summary generation heavily relies on human-engineered features such as manual-labeled summaries [14–17]. Within the gold-standard summary, abstractive model has the copy ability from the original document. Recent researches that use standard transformers to deal with abstractive summarization. The hierarchical structure is to construct the local and global information separately focusing on the semantic features [6,18]. In order to improve the factual consistency of generated summary, Ziqiang et al. [19] propose a fact-aware summarization model with open information extraction and dependency parse technology. Zhu et al. [20] consider the factual relations with graph attention to generating summary.

However, the labeled training data is expensive to acquire. To solve the problem that lacks labeled data, abstractive summarization model mainly focuses on creating a synthetic "review-summary" dataset. Recent researches [16,17] utilize end-to-end, neural network architecture based on autoencoder to perform unsupervised summarization for product reviews without unlabeled data. Firstly, Opinosis [21] is the early abstractive model which applies the structure of graph to eliminate redundancy. Following Opinosis, MeanSum [16] applies the autoencoder model with self-reconstruction loss to learn the feature of reviews and aggregates review features to produce the corresponding summary. Subsequently, Suhara et al. [15] expect that the opinion span can be capable to reconstruct the original review, and establish the sequence-to-sequence samples based on this idea. Following the opinion span, Bražinskas et al. [17] apply the Variation AutoEncoder (VAE) [22] instead of the vanilla autoencoder to facilitate the correlation between summary and reviews. The model achieves a salient performance. Recently, Amplayo et al. [23] conduct the content planning induction to extract the representative review as the pseudo-summary, and reverse to align corresponding reviews. The approach converts the unsupervised scenario to a supervised scenario. Meanwhile, indirect signals (text category and title) are available to select the salient sentences for text extractive summarization as a byproduct of supervised model based on attention mechanism [14]. Similar to their works [14,24], we utilize sequence-to-sequence structure as the basic generated model with weakly human-engineered features. Compared to the above approaches, our proposed model relies on the powerful generator without consideration of hierarchical latent features and explicit summary modeling, shown in Table 1.

Table 1: The characteristic of previous researches

Methods	Advantages	Disadvantages
Opinosis	Opinosis is a flexible framework in that many of its modules can be easily improved or replaced with other suitable implementations.	Since Opinosis is domain-independent and relies on minimal external resources, it cannot group sentences at a deep semantic level.

(Continued)

Table 1: Continued

Methods	Advantages	Disadvantages
OpinionDigest	The model based on sentiment analysis model can filter the selected opinions according to their aspect and/or sentiment.	OpinionDigest may still generate redundant phrases in the summary. The summaries discuss the specified aspect partially, or not at all.
MeanSum	The framework considers the mean of the representations of input reviews. It decodes into a reasonable summary-review while does not rely on any review-specific features.	MeanSum does not provide an unsupervised solution for the more difficult (as there are fewer redundancy cues) single document summarization problem.
CopyCat	The approach controls the "amount of novelty" to go into the new review. The novelty is forced to minimum. The generated text reflects consensus opinions.	Vanilla autoencoder in this approach struggles to properly represent the variety of categories under a single prior. For example, reviews about a sweater can result in a summary about socks.
PlanSum	PlanSum with content planning not only yields an output of higher quality, but also allows the creation of synthetic datasets which are more natural, resembling real-world document-summary pairs.	The approach can not recognize the specific information from reviews. With the length of reviews increasing, the performance may decrease.
Ours	Our proposed model is a simple and weakly supervised framework. The specific information can be retained in generated summary.	The setting of summary structure is considered for different domains.

3 Modeling Approach

In order to satisfy the factual consistency of generated summarization and label the summary for complaint reports, we consider that complaint summary contains two parts: (1) The location, person and organization (named entities) represent the essential and specific information of complaint report. (2) Intention (the existing problem), different from the problem category, is a fine-grained problem. Based on the structure of complaint summary, the combination of entity and intention is proposed to construct the label of complaint summary. Moreover, indirect information is explicitly utilized as the intention of complaint report. The fact corrector and Named Entity Recognition (NER) are utilized to prevent the proposed model generate incorrect information. Therefore, the proposed weakly supervised model consists of four main components: (1) entity extraction from original complaint text, (2) intention generation, (3) summarization generation and (4) fact corrector. The overall architecture of the weakly supervised summarization model is described in Fig.1.

3.1 Problem Definition

Let D denote a set of Chinese complaint reports within an entity set $\left\{e_1, e_2, \ldots, e_{|D|}\right\}$. For each complaint report d, which contains a sequence of words $d = \{w_1, w_2, \ldots, w_n\}$, we define an entity set $\{\mathrm{e}_1, \mathrm{e}_2, \ldots, \mathrm{e}_d\}$. For each entity e_i, entity set contains entity location e_L and entity type e_T. Moreover, we define indirect information $I = \{w_1, w_2, \ldots, w_n\}$ of d, which is shorter than complaint reports.

For each complaint report, our task is to abstractively generate a summary S, which contains the most salient information. Previous abstractive models never explicitly deal with entity and indirect information. On the contrary, we further assume that entity and indirect information can represent a summary of complaint reports. The combination information of complaint report is considered as the core of our framework, as described in the following sections and illustrated in Fig. 1. Finally, the gold-standard summary is not utilized since these do not exist in the most domain.

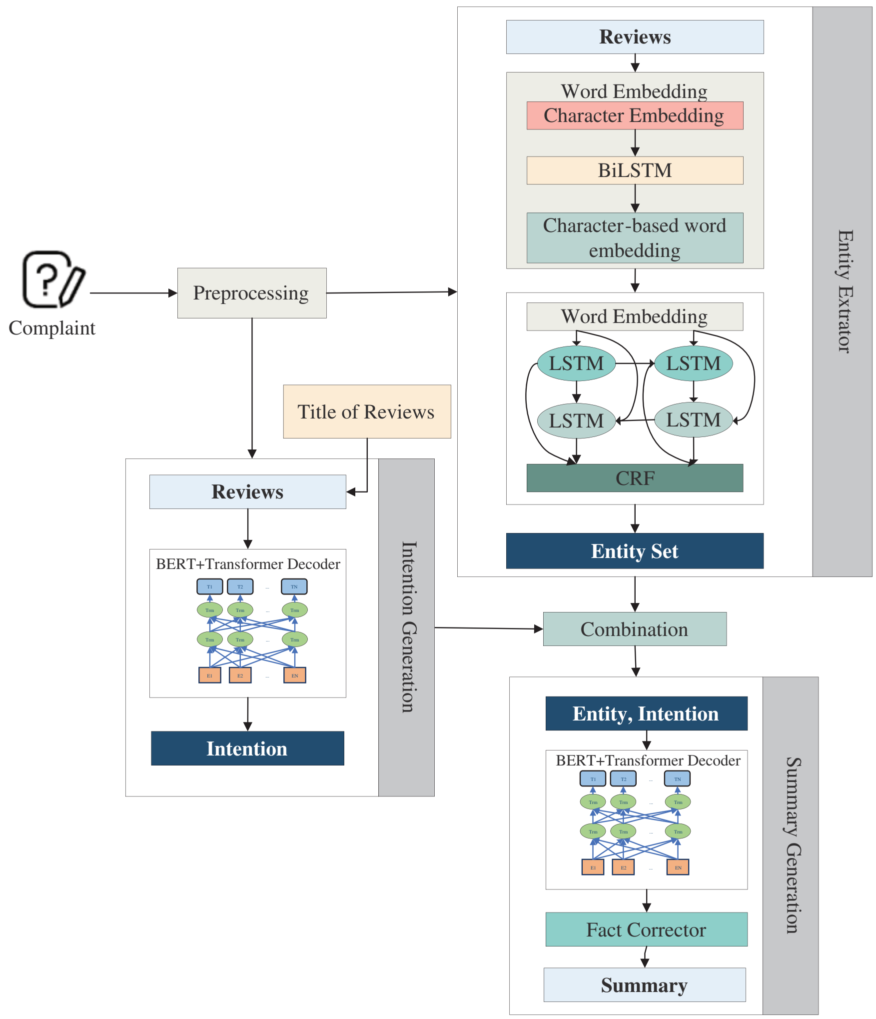


Figure 1: Overall architecture of the proposed model

3.2 Entity Extraction

NER has been studied to extract explicit entities [25–28]. We follow existing models to obtain an entity set E for each review. A pre-trained NER model [29] is utilized to extract entities, shown in Fig. 2. We finetune the pre-trained NER as our entity extractor model. Similar to word embedding of machine translation [30] and text generation [31], entity extraction based on Chinese character embedding can be pre-trained on a large number of Chinese corpora. In entity extraction module, BiLSTM-CRF of pre-trained NER model can adequately represent semantic information for entity extraction. Meanwhile, Bi-directional Long Short-Term Memory (BiLSTM) is to learn character-based word representation. Conditional Random Field (CRF) can capture the dependencies in sequences. After the processing of pre-trained NER, the entity set of each complaint report is generated, which contains three parts: (1) entity name, such as the Shenhai District, Dongfanghejing Community. (2) location of entity and (3) entity types (contain location, person and organization).

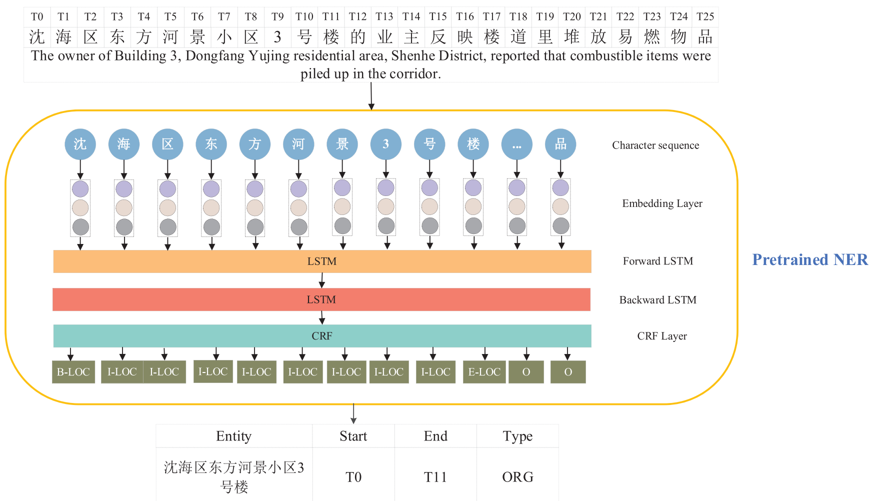


Figure 2: The structure of pretrained NER

3.3 Intention Generation

In order to overcome the deficiency of gold-standard summary, we hypothesize that the indirect information, such as the title, contains the aim of complaint reports through our observation [14]. Since the summary is related to the entity and intention, how to obtain each intention of Chinese complaint report becomes a necessary part. Therefore, we select the 100 titles which describe the intention of urban problems based on each problem category. The sequence-to-sequence model provides us with certain choices in designing our encoder and decoder. The intention generation model extends the recently abstractive summarization of [32], which is a standard Transformer model with BERT as the pre-trained language model.

The semantic features are essential for the representation of Chinese complaint reports. Due to the characteristic of Chinese, we choose the Chinese BERT-base [33] as the pretraining language model to obtain various semantic features. After BERT encoding, the hidden state of the text can be represented as:

$$H_{d_i} = BERTENC\,(d_i) \tag{1}$$

where $BERTENC\,(.)$ is BERT encoding representation with the special label $[cls]$ which represents the sentence embedding vector.

The intention sentence can be synthesized in a sequence of tokens $I_c = (y_1, \ldots, y_m)$. The decoder follows a standard Transformer decoder, which is to generate the output summary by predicting the next token y_m given the previous output hidden state h_{m-1}. It can be represented as:

$$H_m = Trans\left(h_{m-1}, H_{O_i}, v\right) \tag{2}$$

where $Trans\,(.)$ is a Transformer model and v is the fixed length representation.

3.4 Summarization Generation

In order to construct the gold-standard summary for generated model training, we propose a combination strategy in Chinese complaint text. The combination strategy can be described following two steps: (1) entity combination: merge the entities where entity location is continuous and entity type is same to obtain the complete entity. (2) combination of intention and complete entity: insert the complete entities (the location, organization and person) before intention as the gold-standard summary. An example of the combination strategy is shown in Fig. 3.

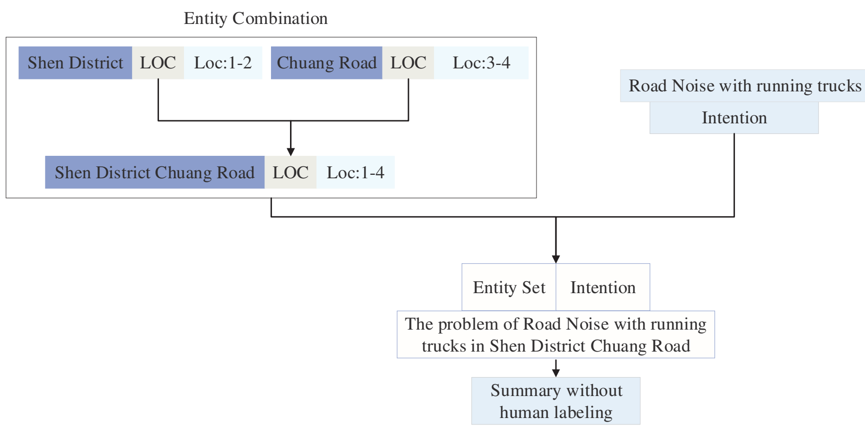


Figure 3: An example of combination processing

After obtaining text-summary pair, a sequence-to-sequence model is utilized to generate the summary S of complaint reports, shown in Fig. 4. The pseudocode of generated strategy is demonstrated in Algorithm 1.

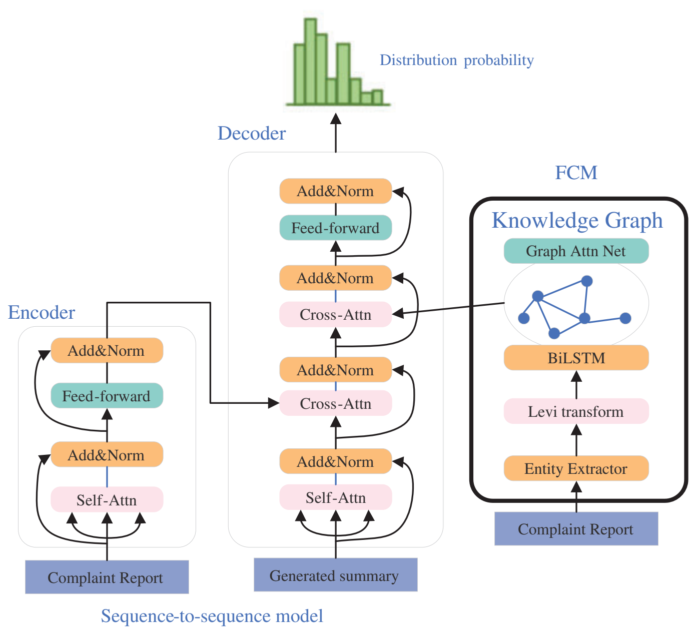


Figure 4: The structure of generated summary model

Algorithm 1: Summarization generation

Output: The generated summary with extracted entity
Input: extracted entity and generated intention
step 1: Entity combination #obtain the complete entity
    if $e_i == e_j$ and $e_{i\_E} + 1 == e_{j\_S}$: #entity type is the same and the start-end is continuous
        $e_i = e_i + e_j$; #generate the complete entity
        $e_{i\_S} = e_{i\_S}$; #set the start location of complete entity
        $e_{i\_E} = e_{j\_E}$; #set the end location of complete entity
step 2: Combination of complete entity and generated intention
    for e_c in complete entity sets: #traverse each entity in the complete entity set
        labeled_summary $= e_c + I_c$; #the combination of the *c-th* entity e_c and complaint intention I_c
step 3: Summary generation #input the complaint reports and labeled_summaries into the seq2seq model
    S_s = word_embedding (source_data);
    S_G = word_embedding (labeled_summary);
    generated_summary = seq2seq (S_s, S_G);

3.5 Fact Corrector

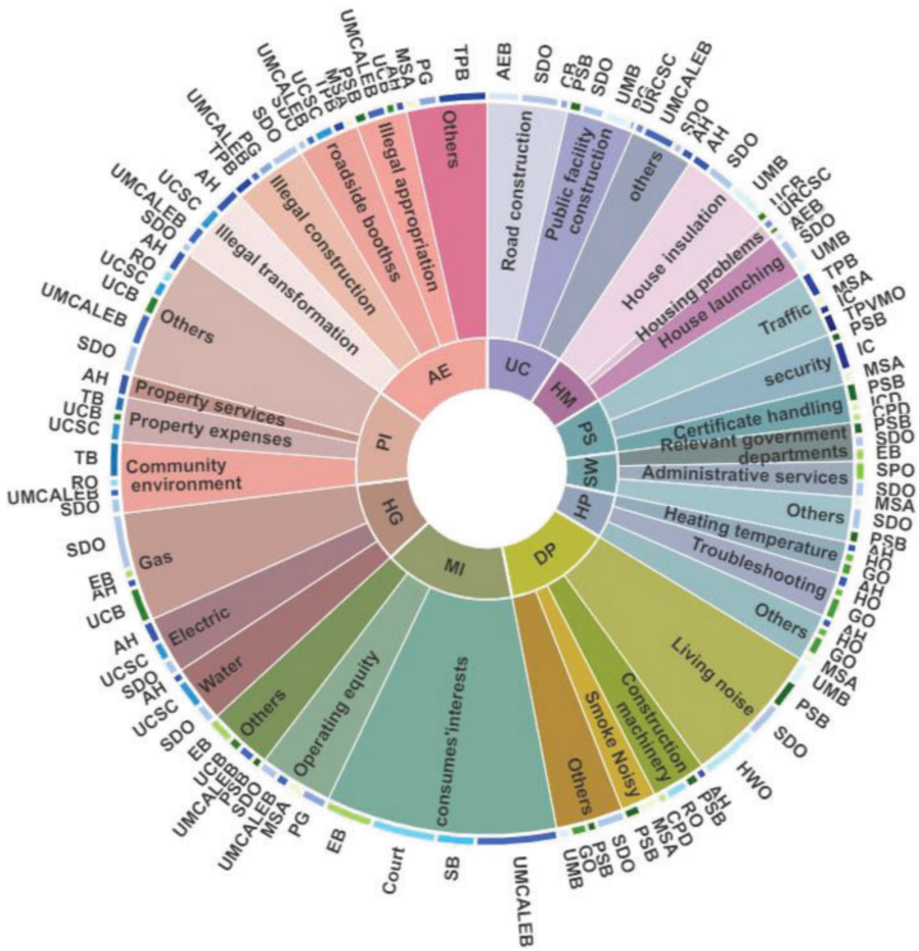
In order to improve the fact corrector of generated summary obtained in Section 3.4, a factual consistency module (FCM) is utilized. FCM is the Transformer model with a graph attention module. The fact corrector is to directly prevent the adaption of pre-trained model and guarantee entity consistency. Similar to the state-of-the-art factual consistency model UniLM [34], parameters of pre-trained model are initialized through BERT-base model. The fine-tuning process is to train a denoising autoencoder. Compared to previous work [19,34], the fact corrector is based on a graph attention mechanism. It can randomly replace with another entity of the same entity type. The summary and complaint report are input into corrector to recover the original summary. The structure of FCM is shown in Fig. 4.

4 Experiments and Results

4.1 Dataset

In order to demonstrate the effectiveness of our model, we use product reviews and complaint reports as our evaluation datasets, covering both general and specific-domain summarization tasks. The complaint records are from 12345-hotline in Shenyang, which are available on the official municipal website (shenyang.gov.cn). The description and length distribution of 12345-hotline complaint reports are shown in Figs. 5 and 6, respectively. It can be observed that the length of complaint report is almost 50–90 tokens(segmented by Chinese character). In Fig. 5, complaint reports contain various problem types and handling departments. A sample of typical complaint records is shown in Table 2.

Moreover, we use business customer reviews-Yelp dataset challenge (Yelp) following [16,35] and the Amazon dataset [36]. Amazon dataset includes 4 categories: electronics, clothing, shoes and jewelry, home and kitchen, health and personal care. There are no gold-standard summaries for two large training corpora, while the small test sets have summaries written by Amazon Mechanical Turk (AMT) crowd-workers.



AEB: Administrative Enforcement Bureau **SDO**: Sub-District Office **PG**: People's Government **Court**: The court
URCSC: Urban and Rural Construction Service Center **TB**: Tax Bureau **IC**: Information Center
UMCALEB: Urban Management Comprehensive Administrative Law Enforcement Bureau **EB**: Education Bureau
AH: Administration of Housing **TPB**: Traffic Police Brigade **UCSC**: Urban Construction Service Center
TPVMO: Traffic Police Vehicle Management Office **MSA**: Market Supervision Administration **HO**: Heating Office
ICD: Immigration Control Division **CPD**: Community Policing Detachment **SPO**: Security Petition Office
GO: Government Office **SB**: Service Bureau **RO**: Resettlement Office **PI**: Property Issues **PS**: Public Security
UCB: Urban Construction Bureau **PSB**: Public Security Bureau **UMB**: Urban Management Bureau
HWO: Health and Wellness Office **UC**: Urban Construction **HM**: House Maintenance **SW**: Service Window
HP: Heating problem **DP**: Disturbing Problems **MI**: Market Issues **HG**: Hydropower gas **PI**: Property Issues
AE: Administrative Enforcement

Figure 5: The distribution of urban problem data

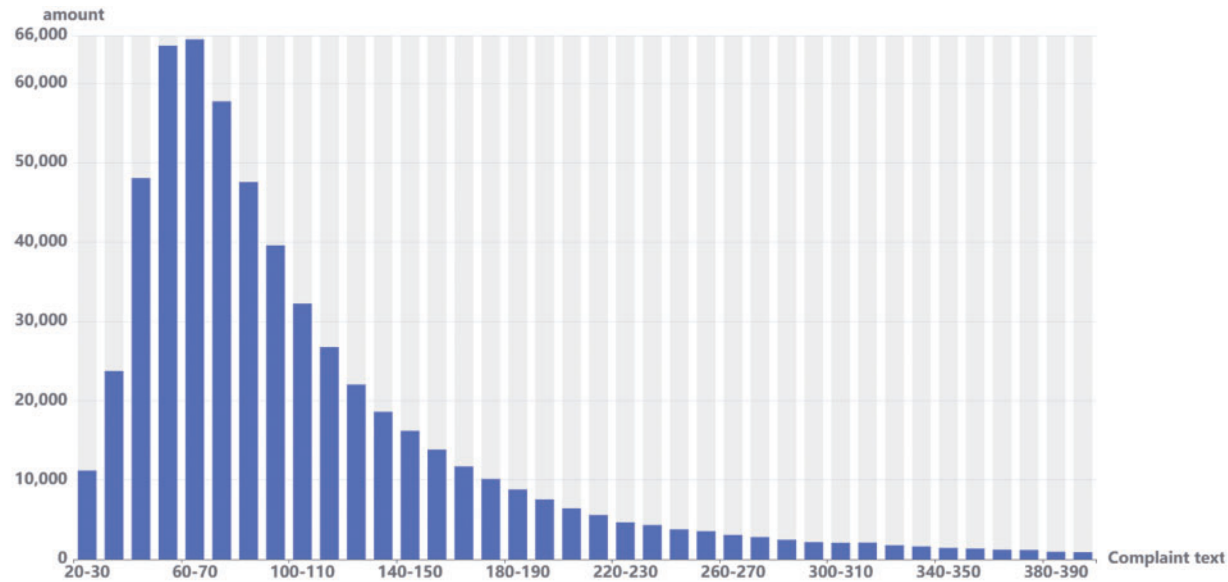


Figure 6: The length of complaint text

Table 2: An example of typical complaint records

Title	Categorization	Complaint record	Handling department
Community greening problem	Urban construction	I am the owner of Baolai Mayflower Phase iii, Shenhai District. The road is being repaired at the west gate of our community. There has a six-meter-wide green belt outside the community. However, the green belt is gone now. There is only a little more than two meters of pavement.	Urban construction bureau
Smell water problem	House maintenance	I am a resident of Unit 3, No. 88-3, Zhulin Road, East District. The smelly water returned from the exterior scene of my unit is unmanaged. There is no property in the community.	Sub-district office

(Continued)

Table 2: Continued

Title	Categorization	Complaint record	Handling department
Community fire escape problem	Public security	Cars have been stopped in the middle of Garden east road influencing cars to move. If an emergency occurs in the neighborhood, ambulance and fire lorry will not be able to enter. It has not been solved by calling 122 two or three times.	Traffic police brigade

4.2 Implementation

For entity extraction, the pre-trained model is trained with the Chinese Wikipedia corpus. For summarization generation, we utilize a standard Transformer encoder [37], pre-initialized with Chinese BERT-base, consisting of 110 M parameters. In addition, in order to keep the same hyperparameters of compared models, we use Stochastic Gradient Descent (SGD) with an initial learning rate of 0.1. Beam size is set as 5. The maximum generation length is 40. In total, we run 500 training epochs for abstractive summarization task. For Yelp and Amazon datasets, an entity represents the same product in business reviews. The pre-trained model is set as BERT-base. The same hyperparameters and parameters are utilized for Yelp and Amazon.

4.3 Results

4.3.1 Automatic Evaluation

We report Recall-Oriented Understudy for Gisting Evaluation (ROUGE) F1 scores [38] as the evaluation scores on Yelp, Amazon and the complaint datasets. ROUGE metric includes ROUGE-1, ROUGE-2, ROUGE-L, etc. ROUGE-1/2 is to count the n-gram (1 or 2) matches between generated summary and gold-standard summary. The gold-standard summary of Yelp and Amazon is the labelled summary in small test sets by AMT while that of Chinese complaint reports is the combination of entity and intention. The overlapped unigram and bigram are reported as the assessing of information. On the contrary, the longest common subsequence is to evaluate the generated summary fluency.

Tables 3 and 4 show the experimental results on the Yelp and Amazon datasets. Compared to Opinosis and OpinionDigest which only consider opinion popularity, our model performs the best. Although our model is not a fully unsupervised framework, labeled data is easier to require by the combination of entity and title in customer reviews. The opinion popularity is essential for the review corpora. Reversely, a simple framework based on sequence-to-sequence model without review popularity can obtain a comparable result. It can infer that our model benefits from the ability of the combination of intention and entity to learn effective language features despite being unsupervised and abstractive. Moreover, although copy-mechanism has been useful in previous summarization models (CopyCat), it is still below the performance of our model. The summary generated by CopyCat is better aligned with reviews, while the detailed information is limited. It can consider that our model keeps the balance of review alignment and detailed information. We then compare our model with state-of-the-art models which lack supervision. It demonstrates that our model obtains an outstanding performance owing to the complaint structure (entity, intention). Although the proposed model falls short of ConsistSum and TransSum that have the advantage of aspect and sentiment, we believe this is because the aspect and sentimental features influence the quality of generated summary while our model does not consider these specific features of reviews.

For Yelp dataset, it can be observed that the result of our proposed model is slightly higher than that of PlanSum in ROUGE-L. Meanwhile, the same situation exists for Amazon dataset, when the proposed model compares to CopyCat model. We infer that the proposed model manages to capture more variety in generating the summary of review corpora.

Table 3: Automatic evaluation of the proposed model on Yelp

Model	Rouge-1	Rouge-2	Rouge-L
LexRank [39]	25.15	2.64	13.37
Opinosis [21]	25.15	2.61	13.54
MeanSum [16]	28.86	3.66	15.91
OpinionDigest [15]	29.30	5.77	18.56
PlanSum [23]	30.14	4.99	17.65
CopyCat [17]	29.47	5.26	18.09
ConsistSum [40]	32.65	7.49	20.87
sequence-to-sequence + NER	30.96	7.42	18.67

Table 4: Automatic evaluation of the proposed model on Amazon

Model	Rouge-1	Rouge-2	Rouge-L
LexRank [39]	28.74	5.47	16.75
Opinosis [21]	28.42	4.57	15.50
MeanSum [16]	29.20	4.70	18.15
OpinionDigest [15]	–	–	–
PlanSum [23]	32.87	6.12	19.05
CopyCat [17]	31.97	5.81	20.16
ConsistSum [40]	33.32	5.94	21.41
TranSum [35]	34.23	7.24	20.49
sequence-to-sequence + NER	33.27	6.95	21.08

Note: *Our model results are in bold. "-" is to indicate unreported results or unfound outputs.

Table 5 reports ROUGE scores in 12345-hotline complaint text. Our model obtains a significant improvement in ROUGE scores. We find that our model with entity set improves ROUGE-L scores by more than 16 compared with LexRank. It can demonstrate that the entity-based generated model substantially improves the coherence and reduces redundancies in generated summary. Moreover, the scores of Rouge-1/2 improve 15 and 14. The results confirm that our model is suitable for complaint text although it is a simple model. We can consider that the structure of proposed model captures the various semantic features.

Table 5: Automatic evaluation of proposed model on complaint dataset

Model	Rouge-1	Rouge-2	Rouge-L
LexRank	40.13	30.74	38.31
Seq2seq	46.68	23.55	42.33
Seq2seq + NER	55.21	44.74	54.35

Although the difference exists in Chinese complaint reports and English reviews, the above result can be proved that the proposed model can be utilized in Chinese or English corpora. We infer that the different pre-trained language model plays the vital role in summarization model. In general, the proposed framework can obtain a comparable result in Yelp and Amazon dataset to some extent. These results can demonstrate that the proposed model is effective for real application. Moreover, for Chinese complaint reports, the proposed framework achieves an obviously success of the generated summary. It can be proved that the generated summary of the proposed model is coherent and contains the salient information.

4.3.2 Human Evaluation

In addition to automatic evaluation, we also assess generated summaries by eliciting human judgments. Informativeness, coherence and redundancy are utilized to evaluate the quality of generated summary by 3 urban problem administrators [17,23,41]. The 3 user criteria are illustrated as follows: Informativeness: salient information (concern complaint intention, location, or organization) exists in the generated summary. Coherence: summary is coherent and easy to read. Redundancy: summary is avoided by the overlapped information and has less redundancy.

The framework's ability to generate a summary with NER is shown in Table 6. Firstly, compared to Lexrank, our proposed model obtains a promising result. The result of standard sequence-to-sequence model is lower than LexRank without NER. Since LexRank extracts several sentences, the process ensures the informativeness of generated summary. However, the redundancy and coherence scores of LexRank are lower than those of sequence-to-sequence because of the ability to generate new tokens. The generated summary of sequence-to-sequence is better compressed than that of LexRank. Moreover, the improvement of sequence-to-sequence model with NER is higher than the improvement of LexRank with NER. It can demonstrate that NER module effectively helps the performance of proposed model. Finally, with the fact corrector, the sequence-to-sequence model generates summaries in a more concise language. It can infer that the fact corrector module can improve the faithfulness, conciseness and readability of generated summary.

4.3.3 Case Study

In order to demonstrate the factual consistency of summary generated by our model, a case study is shown in Table 7. Colors show the same entity type between complaint text and generated summary. The underlined text represents the incorrect information in the generated summary. As shown in Table 7, the traditional sequence-to-sequence model generates the summary with incorrect entities or without entities. On the contrary, the generated summary of proposed model contains the correct entities and intention. The result demonstrates: (1) the proposed model combining entity extraction and intention can generate meaningful summaries from thousands of urban problem complaint data and (2) entity extraction indirectly produces control-label summaries in terms of factual consistency.

Table 6: The human evaluation result of the proposed model on complaint reports

Model	I-score	C-score	R-Score
LexRank	1.338	2.397	1.387
+NER	1.438	2.461	1.573
+ Fact corrector	1.546	2.525	1.592
Sequence to sequence	1.232	2.412	1.446
+NER	1.644	2.814	1.859
+ Fact corrector (our)	1.682	2.956	1.921

Table 7: case study of the proposed abstractive summarization model

Entity	Complaint	Summary wo entity	Summary w NER
Entity with location	I reflect that LOC A has started heating on December 1 and the temperature is nearly 15 degrees. The heating company is ORG A. We have reported this matter to the heating company many times but did not solve the problem. Please investigate and deal with the relevant departments.	The problem that heating does not meet the standard	The problem that heating does not meet the standard in LOC A.
Entity with organization	I had reported that LOC A is a fire escape occupied by the two residential properties. A property puts stone bent down and rents the land to a second-hand car company. I want to complain to property ORG A that occupies the fire escape many times.	The problem of ORG B property occu-pying a LOC B fire escape.	The problem of property ORG A occupying the fire escape.
Entity with person	I am a resident of LOC A. I have paid a lump sum for the new rural insurance scheme from 2011 to 2020. It was given in cash without any receipts to PER A, a female worker in LOC A. However, my account only showed my expenses. I hope relevant departments can verify and deal with it.	The problem that ORG A does not act.	The problem that PER A does not act.

(Continued)

Table 7: Continued

Entity	Complaint	Summary wo entity	Summary w NER
Entity with all	My wife PER A is 60 years old, living in LOC A. PER A has been paid new farming and social security in August. PER A began to deal with retirement. PER A should take 1221 yuan fee per month. However, until now, we have not received the retirement salary for nearly 2 months. The office of ORG A tells me to wait until the end of September. Please solve it as soon as possible	The problem of ORG A retirement issues.	The problem of ORG A PER A retir-ement issues in LOC A.

The first three examples in Table 7 illustrate that only one type of entity supplement generates the abstractive summary with inadequate information. It can be observed that only one type of entity combination lacks salient information concerning entities. Moreover, the last example is the proposed model with all entity types. It can be shown that our model performs robustly even for numerous complaint reports. Since the proposed model considers the structure of complaint summary (location, person, intention, etc.), the quality of generated summary is not affected by the length or the entity number of complaint text. We can consider that the generated summary expresses a better-informed ability. Meanwhile, it can be proved that the proposed model provides a handling processing to quickly solve the urban problem for government.

5 Conclusion and Future Work

We describe a simple yet powerful framework with a weakly supervised model to solve the problems of factual consistency and label absence for complaint abstractive summarization. Our proposed model is an entity-based abstractive summarization. It does not heavily rely on gold-standard summaries. Evaluation experiments on Yelp, Amazon and complaint reports demonstrate that our proposed framework outperforms other state-of-the-art unsupervised summarization approaches. It can be proved that the proposed summarization framework can help in dealing with complaint reports and effectively reducing the manual cost of city development for government department.

Text-To-Text Transfer Transformer (T5) can solve a variety of tasks as simple text-to-text mapping problems. It is considered as the further language model. Moreover, the proposed model aims to construct a pseudo text-summary pair without any fact reasoning. Therefore, fact reasoning should be investigated. Finally, the other indirect information can be studied as the intention of complaint reports.

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

Funding Statement: This work is partially supported by National Natural Science Foundation of China (62276058, 61902057, 41774063), Fundamental Research Funds for the Central Universities (N2217003), Joint Fund of Science & Technology Department of Liaoning Province and State Key Laboratory of Robotics, China (2020-KF-12-11).

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

References

[1] Y. Zheng, T. Liu, Y. L. Wang, Y. M. Zhu, Y. C. Liu *et al.,* "Diagnosing New York city's noises with ubiquitous data ubiquitous computing," in *Proc. UbiComp*, Seattle, WA, USA, pp. 715–725, 2014.
[2] L. Zhu, B. Zhou, C. Wu and J. Zhou, "Investigation to the intelligence of urban management under the new situation," *Science & Technology Association Forum*, vol. 2, pp. 154–155, 2013.
[3] C. Kaushal and D. Koundal, "Recent trends in big data using hadoop," *Int. Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 8, no. 1, pp. 39–49, 2019.
[4] J. Liu, "Harvesting and summarizing user-generated content for advanced speech-based human-computer interaction," Doctoral Dissertation, Massachusetts Institute of Technology, 2012.
[5] V. B. Raut and D. D. Londhe, "Opinion mining and summarization of hotel reviews," in *Int. Conf. on Computational Intelligence and Communication Networks*, Bhopal, India, pp. 100–115, 2014.
[6] Y. Liu and M. Lapata, "Hierarchical transformers for multi-document summarization," in *Proc. ACL*, Florence, Italy, pp. 5070–5081, 2019.
[7] T. Cai, M. Shen, H. Peng, L. Jiang and Q. Dai, "Improving transformer with sequential context representations for abstractive text summarization," *Natural Language Processing & Chinese Computing*, vol. 11838, pp. 512–524, 2019.
[8] E. Linhares, S. Huet, J. M. Torres and A. C. Linhares, "Compressive approaches for cross-language multi-document summarization," *Data & Knowledge Engineering*, vol. 125, no. 10, pp. 101763, 2020.
[9] A. M. Rush, S. Chopra and J. Weston, "A neural attention model for abstractive sentence summarization," in *Proc.EMNLP*, Lisbon, Portugal, pp. 379–389, 2015.
[10] M. A. Ranzato, S. Chopra, M. Auli and W. Zaremba, "Sequence level training with recurrent neural networks," in *Proc. ICLR*, Puerto Rico, pp. 1–16, 2016.
[11] J. Li, M. T. Luong and D. Jurafsky, "A hierarchical neural autoencoder for paragraphs and documents," in *Proc. ACL-IJCNLP*, Beijing, China, pp. 1106–1115, 2015.
[12] A. Jadhav and V. Rajan, "Extractive summarization with swap-net: Sentences and words from alternating pointer networks," in *Proc. ACL*, Melbourne, Australia, pp. 142–151, 2018.
[13] J. Tan, X. Wan and J. Xiao, "Abstractive document summarization with a graph-based attentional neural model," in *Proc. ACL*, Vancouver, Canada, pp. 1171–1181, 2017.
[14] Y. Zhuang, Y. Lu and S. Wang, "Weakly supervised extractive summarization with attention," in *Proc. SIGDIAL*, Singapore, pp. 520–529, 2021.
[15] Y. Suhara, X. Wang, S. Angelidis and W. C. Tan, "OpinionDigest: A simple framework for opinion summarization," in *Proc. ACL*, Online, Seattle, Washington, United States, pp. 5789–5798, 2020.
[16] E. Chu and P. Liu, "Meansum: A neural model for unsupervised multi-document abstractive summarization," in *Proc. ICML*, Long Beach, CA, USA, pp. 1223–1232, 2019.
[17] A. Bražinskas, M. Lapata and I. Titov, "Unsupervised opinion summarization as copycat-review generation," in *Proc. ACL*, Online, Seattle, Washington, United States, pp. 5151–5169, 2020.
[18] L. Yang and L. Mirella, "Learning structured text representations," *Transactions of the Association for Computational Linguistics*, vol. 6, no. 1, pp. 63–75, 2018.
[19] C. Ziqiang, W. Furu, L. Wenjie and L. Sujian, "Faithful to the original: Fact aware neural abstractive summarization," in *Proc. AAAI*, Louisiana, USA, pp. 4784–4791, 2018.
[20] C. Zhu, W. Hinthorn, R. Xu, Q. Zeng, M. Zeng *et al.,* "Enhancing factual consistency of abstractive summarization," in *Proc. NAACL*, Online, Seattle, Washington, United States, pp. 718–733, 2021.
[21] K. Ganesan, C. Zhai and J. Han, "Opinosis: A graph-based approach to abstractive summarization of highly redundant opinions," in *Proc. ICCL*, Beijing, China, pp. 340–348, 2010.
[22] F. Xianghua, W. Yanzhi, X. Fan, W. Ting, L. Yu *et al.,* "Semi-supervised aspect-level sentiment classification model based on variational autoencoder," *Knowledge-Based Systems*, vol. 171, pp. 81–92, 2019.

[23] R. K. Amplayo, S. Angelidis and M. Lapata, “Unsupervised opinion summarization with content planning,” in *Proc. AAAI*, Online, vol. 35, pp. 12489–12497, 2021.
[24] G. D. Fabbrizio, A. Stent and R. Gaizauskas, “A hybrid approach to multi-document summarization of opinions in reviews,” in *Proc.INGL*, Philadelphia, Pennsylvania, U.S.A, pp. 54–63, 2014.
[25] C. Dong, J. Zhang, C. Zong, M. Hattori and H. Di, “Character-based LSTM-CRF with radical-level features for Chinese named entity recognition,” *Natural Language Understanding & Intelligent Applications*, vol. 10102, pp. 239–250, 2016.
[26] X. Cai, S. Dong and J. Hu, “A deep learning model incorporating part of speech and self-matching attention for named entity recognition of Chinese electronic medical records,” *BMC Medical Informatics & Decision Making*, vol. 19, no. 2, pp. 101–109, 2019.
[27] G. Wu, G. Tang, Z. Zhang Z.Wang and Z. Wang, “An attention-based BiLSTM-CRF model for Chinese clinic named entity recognition,” *IEEE Access*, vol. 7, pp. 113942–113949, 2019.
[28] L. Li, J. Zhao, L. Hou, Y. Zhai, J. Shi *et al.,* “An attention-based deep learning model for clinical named entity recognition of Chinese electronic medical records,” *BMC Medical Informatics & Decision Making*, vol. 19, no. 5, pp. 1–11, 2019.
[29] R. Panchendrarajan and A. Amaresan, “Bidirectional LSTM-CRF for named entity recognition,” in *Proc. PACLIC*, Hong Kong, China, pp. 531–540, 2018.
[30] S. M. Singh and T. D. Singh, “An empirical study of low-resource neural machine translation of manipuri in multilingual settings,” *Neural Computing and Applications*, vol. 34, pp. 1–22, 2022.
[31] U. Akujuobi, M. Spranger, S. K. Palaniappan and X. Zhang, “T-PAIR: Temporal node-pair embedding for automatic biomedical hypothesis generation,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 6, pp. 2988–3001, 2022.
[32] Y. Liu and M. Lapata, “Text summarization with pretrained encoders,” in *Proc. EMNLP-IJCNLP*, Hong Kong, China, pp. 3730–3740, 2019.
[33] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proc.ACL*, Minneapolis, MN, USA, pp. 4171–4186, 2019.
[34] L. Dong, N. Yang, W. Wang, F. Wei, X. Liu *et al.,* “Unified language model pre-training for natural language understanding and generation,” in *Proc. NeurIPS*, Vancouver, Canada, pp. 13063–13075, 2019.
[35] K. Wang and X. Wan, “TransSum: Translating aspect and sentiment embeddings for self-supervised opinion summarization,” in *Proc.ACL*, Punta Cana, Dominican Republic, pp. 729–742, 2021.
[36] R. He and J. McAuley, “Ups and downs: modeling the visual evolution of fashion trends with one-class collaborative filtering,” in *Proc. of the 25th Int. Conf. on World Wide Web*, Montréal Québec Canada, pp. 507–517, 2016.
[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones *et al.,* “Attention is all you need,” *in Proc. NIPS*, Long Beach, California, pp. 6000–6010, 2017.
[38] C. Y. Lin, “Rouge: A package for automatic evaluation of summaries,” in *Proceedings of Text Summarization Branches out*, Barcelona, Spain, pp. 74–81, 2004.
[39] G. Erkan and D. R. Radev, “Lexrank: Graph-based lexical centrality as salience in text summarization,” *Journal of Artificial Intelligence Research*, vol. 22, pp. 457–479, 2004.
[40] W. Ke, J. Gao, H. Shen and C. Xueqi, “ConsistSum: unsupervised opinion summarization with the consistency of aspect, Sentiment and Semantic,” in *Proc. WSDM*, NY, US, pp. 467–475, 2022.
[41] G. Bao and Y. Zhang, “Contextualized rewriting for text summarization,” *Proc. AAAI*, vol. 35, no. 14, pp. 12544–12553, 2021.