Pattern Recognition 146 (2024) 109952

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A Cross-Lingual Summarization method based on cross-lingual Fact-relationship Graph Generation☆

Yongbing Zhang, Shengxiang Gao, Yuxin Huang, Kaiwen Tan, Zhengtao Yu *

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, 650500, PR China
Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming, 650500, PR China

ARTICLE INFO

Keywords:
Cross-lingual summarization
Fact-relationship graph
Deliberation network
Graph generation
Factual inconsistency

ABSTRACT

The aim of cross-lingual summarization (CLS) is to condense the content of a document in one language into a summary in another language. In essence, a CLS model requires both translation and summarization capabilities, which presents a unique challenge, as the model must effectively tackle the difficulties associated with both tasks simultaneously (e.g., semantic alignment, information compression and factual inconsistency). Graph-based semantic representation can model important text information in a structured manner, which may alleviate these challenges. Therefore, in this paper, we propose a Cross-Lingual Summarization method based on cross-lingual Fact-relationship Graph Generation (FGGCLS). Specifically, we first construct fact-relationship graphs for source language documents and target language summaries. Then, we introduce a cross-lingual fact-relationship graph generation method, which converts the CLS problem into a cross-lingual fact-relationship graph generation problem. This approach simplifies semantic alignment and information compression through the generation of graphs and leads to improved fact consistency. Finally, the generated fact-relationship graph of the target language summary serves as a draft for generating the summary, which enhances the quality of the generated summary. We conduct systematic experiments on the Zh2EnSum and En2ZhSum datasets, and the results demonstrate that our method can effectively improve the performance of CLS and alleviate factual inconsistency.

1. Introduction

Cross-lingual summarization (CLS) is a natural language processing task that generates concise summaries in a target language (e.g., Chinese) from a source language (e.g., English) while maintaining the key information and context. The primary goal of CLS is to enable quick comprehension of documents in unfamiliar languages, thereby improving information acquisition efficiency. In recent years, CLS has attracted significant research interest; however, it remains an extremely challenging task, as it requires both translation and summarization capabilities. Therefore, challenges such as cross-lingual semantic alignment, information compression, and factual inconsistency pose significant obstacles in generating precise and coherent cross-lingual summaries.

Recently, many researchers have utilized multi-task [1–5] and knowledge distillation [6,7] methods to solve the challenges of CLS. The multi-task methods incorporate related tasks such as machine translation and monolingual summarization into the training process, thus enabling multi-task learning to enhance the quality of the summary. The knowledge distillation methods, on the other hand, use two models of machine translation and monolingual summarization as teacher models to impart the learned knowledge to the student model to improve CLS performance. However, these methods have some limitations. First, they heavily rely on additional datasets of large-scale machine translation and monolingual summarization. Second, they require simultaneous training for CLS and monolingual summarization or alternating between training for CLS and machine translation. This training process can be quite time-consuming. Last, they do not explicitly model the process of cross-lingual and key information compression, which may lead to factual inconsistency issues, and the model may not be interpretable.

☆ This work was supported by the National Natural Science Foundation of China [grant numbers U21B2027, 61972186, 62266027, 62266028]; Yunnan Provincial Major Science and Technology Special Plan Projects, China [grant numbers 202103AA080015, 202202AD080003]; General Projects of Basic Research in Yunnan Province, China [grant numbers 202201AT070915, 202201AT070768]; Kunming University of Science and Technology, China "double first-class" joint project [202201BE070001-021].

* Corresponding author at: Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, 650500, PR China.
E-mail addresses: zhangyongbing419@163.com (Y. Zhang), ztyu@hotmail.com (Z. Yu).

https://doi.org/10.1016/j.patcog.2023.109952
Received 30 March 2023; Received in revised form 3 September 2023; Accepted 5 September 2023
Available online 9 September 2023

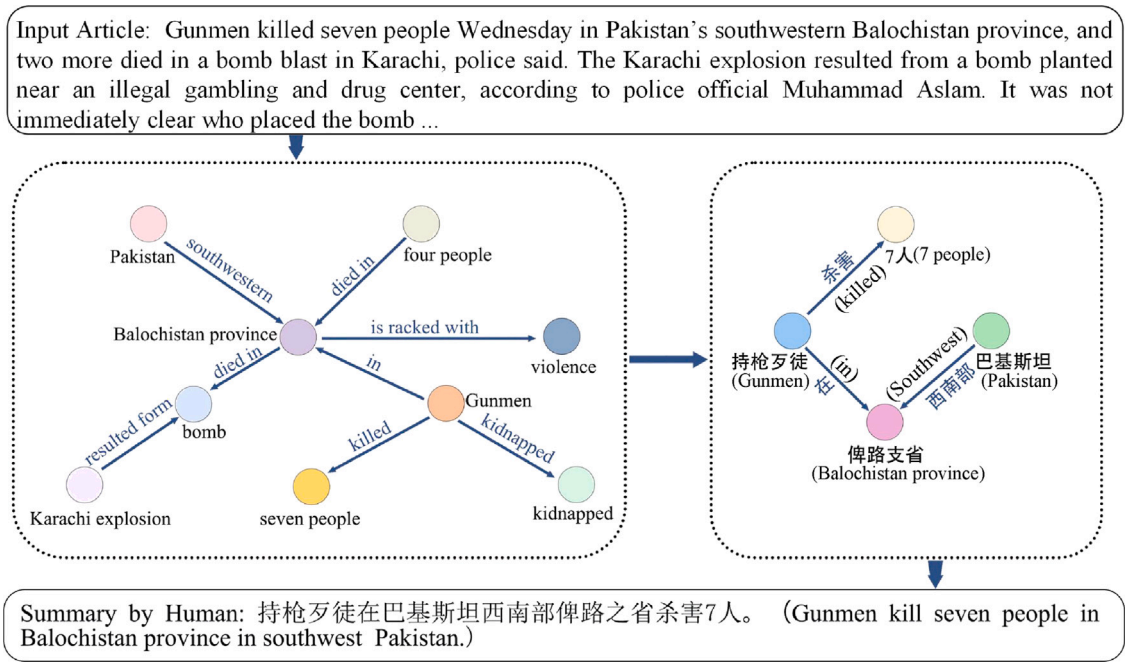

Fig. 1. A random sample (i.e. a pair of articles and summaries) of the En2ZhSum training dataset. The English graph structure represents the fact-relationship graph extracted from the source language article, while the Chinese graph structure represents the fact-relationship graph extracted from the target language summary.

Graph-based semantic representation can model important text information in a structured manner, which may alleviate these challenges. Therefore, we conducted a case study on a random sample of the En2ZhSum training dataset to validate this idea. We found that the source language document and target language reference summary contain a significant amount of fact triple information that succinctly captures the key information of the documents and summaries and is relevant. As shown in Fig. 1, the input source language documents include fact triples such as (Pakistan, southwestern, Balochistan province), (Gunmen, killed, seven people), and (Balochistan province, is racked, violence), which enable us to extract crucial information, for instance, "Balochistan province is located in the southwestern region of Pakistan" and "an incident of shooting took place in Balochistan province". Similarly, the target language summary also contains analogous information in the corresponding triples, such as "(巴基斯坦, 西南部, 俾路支省)" (i.e. Pakistan, southwestern, Balochistan province) and "(持枪歹徒, 杀害, 7人)" (i.e. Gunmen, killed, seven people). The triple information referenced in the summary of the target language is present in the source language document, albeit expressed using a different language. Graphs are effective in organizing and modelling multiple triple relationships in a structured manner, so they can be used to help information compress and enhance factual consistency. Additionally, as a mathematical concept and data structure, graphs are inherently language-independent, which can effectively reduce the difficulty of semantic alignment.

Based on these observations, we propose a **C**ross-**L**ingual **S**ummarization method based on cross-lingual **F**act-relationship **G**raph **G**eneration (FGGCLS). This approach addresses the challenges of CLS by converting it into a cross-lingual graph generation task. First, we extract fact triple information from source language documents and target language reference summaries and construct their fact-relationship graphs. Then, we explicitly associate crucial fact information in the documents and summaries by mapping the source language fact-relationship graph into target language fact-relationship graphs. This method alleviates the semantic alignment and information compression challenges in traditional CLS, thereby improving fact consistency. Finally, the target language summary graph serves as a draft, and a more precise summary text is generated under its guidance. Overall, FGGCLS belongs to the application of graph-based pattern recognition. It utilizes a complex graph to model fact-relationships within a document and bridge fact consistency between source language documents and target language summaries.

Experiments conducted on the Zh2EnSum and En2ZhSum datasets demonstrate that the proposed FGGCLS significantly outperforms baseline models; this indicates that cross-lingual fact-relationship graph generation plays a crucial role in improving the performance of CLS. Moreover, extensive human evaluation experiments verify that the generated target language summaries are more factually consistent and trustworthy.

Our research makes several contributions to the field of graph machine learning and CLS:

(1) We propose a novel method for CLS, which transforms the problem of CLS into a structured graph generation task. By explicitly modelling important fact information from source language documents and target language reference summaries, this method alleviates the challenge of semantic alignment and information compression.
(2) We use the generated target language fact-relationship graph as a draft to guide the generation of a more accurate and reliable target language summary.
(3) Our approach for CLS solely depends on the dataset itself, without requiring large-scale parallel datasets of machine translation, which significantly reduces the data requirements of the model.
(4) Our experimental results demonstrate the effectiveness of our proposed model on both the Zh2EnSum and En2ZhSum datasets and reflect significant improvements over baseline models in terms of summarization quality.

2. Related work

The traditional CLS method has mainly adopted the pipeline approach, which decomposes the task into two subtasks: translation and summarization. This type of approach can be implemented in two ways: translate-then-summarize [8–13] or summarize-then-translate [14–16]. In the translate-then-summarize approach, researchers have explored various techniques for generating English summaries from documents

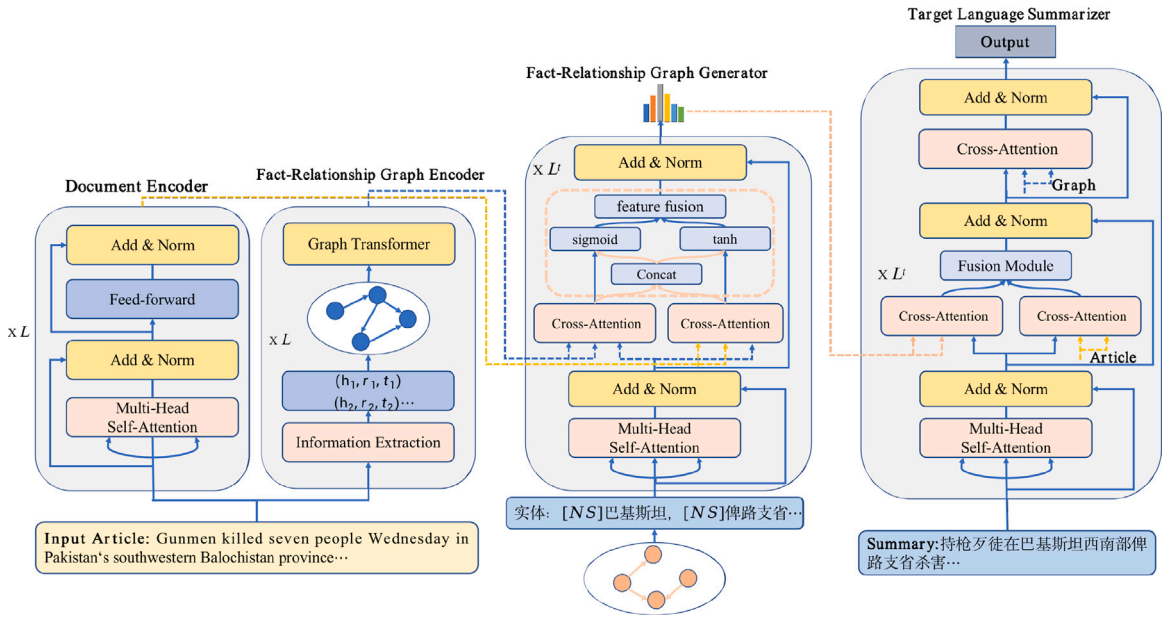

Fig. 2. The FGGCLS framework consists of four components: **Fact-Relationship Graph Encoder**, which models complex factual relationships within the source language document, captures both global document structure information and semantic representation. **Document Encoder**, which encodes the entire document and captures the local semantic information of the source language document. **Fact-Relationship Graph Generator**, which generates the fact-relationship graph of the target language based on the hidden state representation of the encoder. **Target Language Summarizer**, which aims to generate summaries based on the representation of the encoder and the first-stage target language fact-relationship graph. The summary is generated through deliberation.

in other languages. For instance, Leuski et al. [17] translated Hindi documents into English before generating captions. Ouyang et al. [13] developed a robust abstract summarization system that can generate fluent summaries from machine-translated documents, using both noisy English documents and clean English reference summaries for training. Wan [9] and Boudin et al. [8] employed bilingual feature information to predict the translation quality of each sentence in the source language document before generating the final summary via a ranking algorithm. Yao et al. [18] proposed a compression method that calculated sentence scores based on aligned bilingual phrases obtained from machine translation services and removed redundant or poorly translated phrases to perform compression. Linhares Pontes et al. [11] considered bilingual vocabulary chunks when computing sentence similarity and further compressed sentences at both the single-sentence and multi-sentence levels. In the summarize-then-translate approach, some researchers first generate source language summaries and then translate them into the target language. Orăsan and Chiorean [14] proposed using the maximum marginal relevance (MMR) algorithm to generate Romanian news summaries, which were then translated into English via machine translation. Wan et al. [15] employed an SVM model to predict the translation quality of each English sentence, selecting high-quality and informative sentences to form a summary, which was then translated into Chinese using Google Translate. While such pipeline approaches are intuitive, they can also encounter issues such as error propagation, the need for a large corpus or expensive training of the translation model, and delays in the inference process.

In recent years, there has been a gradual shift from pipeline-based methods to end-to-end model architectures in CLS, especially since Zhu et al. [1] proposed a large-scale cross-lingual summary dataset. However, using the end-to-end model directly in CLS still poses challenges, as it requires the model to have both translation and summarization capabilities. To overcome this, several studies have combined related tasks such as translation and summarization with CLS to train a unified CLS model that can benefit from related tasks. Zhu et al. [1] proposed incorporating monolingual summarization or machine translation into CLS training by using a shared Transformer encoder to encode the input sequence for CLS or related tasks (monolingual summarization or machine translation), and two independent Transformer decoders to perform CLS and related task decoding. Cao et al. [3] proposed utilizing two encoders and two decoders to jointly learn cross-lingual alignment and abstract summarization. The MCLAS model was proposed by Bai et al. [4], which uses a unified decoder to generate monolingual and cross-lingual summaries sequentially and shares the decoder to learn cross-lingual alignment and interaction of abstract summaries simultaneously. Liang et al. [5] used conditional variational autoencoders (CVAE) to capture the hierarchical relationship between related tasks and CLS. Moreover, some studies have used translation/summarization tasks as teacher models to teach CLS models; this enables the CLS model to learn not only from the label but also from the output and hidden state of the teacher model. Shen et al. [6], Duan et al. [19], Nguyen and Luu [7] confirmed that the knowledge distillation framework can improve the performance of CLS. However, these methods do not explicitly model the process of cross-lingual and important information compression, and their model training relies on large-scale machine translation or monolingual summarization dataset support.

Graphs, as non-Euclidean data structures, have evolved into diverse forms, including heterogeneous graphs [20], and have been successfully applied to various fields, such as community detection [21]. Graph can effectively organize and model the crucial information within text and present this information in a structured and easily digestible format [22,23]. For example, Yu et al. [24] proposed AS-GCN, which unified neural topic model and graph convolutional networks, and can capture both the local word-sequence semantic structure and the global topic semantic structure. Therefore, several studies have incorporated graphs into text summarization techniques to improve the model's capacity for identifying and summarizing critical information. For example, Fernandes et al. [25] employed a graph neural network to improve token-level entity type information by utilizing a sequence encoder. Fan et al. [26] demonstrated the effectiveness of encoding a linearized knowledge graph obtained from OpenIE in multi-document summarization tasks. It is worth noting that OpenIE and LTP are often used to extract entities and entity relationships, and then a graph can be constructed with entities as nodes and relationships as edges. Koncel-Kedziorski et al. [27] introduced a novel graph transformer encoder that utilizes the relational structure of a knowledge graph to generate text, which was successfully applied to scientific text summarization.

In cross-lingual tasks, Jiang et al. [28] utilized TextRank to extract crucial cues from input sequences and then constructed article graphs based on these cues; they subsequently encoded both the cues and the article graphs using clue encoders and graph encoders, respectively. Finally, during decoding, both sequence encoding information and graph encoding information were considered to generate the final summary. However, while these methods utilized graphs to enhance the representation of key information in input documents, they did not consider how to use the graphs to solve the main challenges of CLS.

3. Background

3.1. Cross-lingual summarization

Recently, Zhu et al. [1] introduced the application of the transformer model for CLS. The transformer model is composed of multiple layers of stacked encoders and decoders. In our study, we adopt this model framework to implement our approach. First, the encoder models the source language document, represented as $X = (x_1, x_2, \dots, x_m)$, into a continuous vector $Z = (z_1, z_2, \dots, z_m)$. Next, the decoder generates the target language summary sequence, represented as $Y = (y_1, y_2, \dots, y_n)$, based on Z. The training objectives of the CLS model are defined as follows:

$$L_\theta = \sum_{t=1}^{N} log P(y_t | y_{<t}, X; \theta) \tag{1}$$

Here, θ is the parameter for model training and $y_{<t}$ is a partial summary of the target language reference.

3.2. Deliberation network

Inspired by the iterative refinement process often employed in human translation and article writing, Xia et al. [29] integrated the deliberation process into the encoder–decoder framework, which effectively enhanced the overall quality of the generated sequences. The deliberation network is comprises a two-stage decoder. The first-stage decoder is responsible for decoding and generating the original sequence information, while the second-stage decoder refines and improves the output based on the decoding results of the first stage. Specifically, given an input sequence $X = (x_1, x_2, \dots, x_m)$, the first-stage decoder of the deliberation network model decodes the original sequence information $Y' = (y'_1, y'_2, \dots, y'_l)$. Subsequently, the second-stage decoder conducts a second round of deliberation on the output sequence generated by the first stage, resulting in the final output sequence $Y = (y_1, y_2, \dots, y_n)$. The training objectives of the deliberation network are as follows:

$$L(\theta_e, \theta_1, \theta_2) = \sum_{y' \in Y} P(y' | E(X; \theta_e); \theta_1) \cdot log P(y | y', E(X; \theta_e); \theta_2) \tag{2}$$

Where θ_e refers to the trainable parameters of the encoder, $\theta_1 and \theta_2$ refer to the training parameters of the first-stage and second-stage decoders, respectively.

4. Method

4.1. Problem definition

We denote a parallel CLS data pair as $D = (X, Y)$, where $X = (x_1, x_2, \dots, x_m)$ refers to the source language document input to the model, and $Y = (y_1, y_2, \dots, y_n)$ represents the target language reference summary. The token lengths of the source language document and target language reference summary are denoted by m and n, respectively. Given a particular source language document X, our model produces $\hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{\hat{n}})$.

Furthermore, to enhance the performance of our CLS model, we propose the idea of cross-lingual fact-relationship graph generation. The source language fact-relationship graph is denoted as $F^s = (H^s, R^s, T^s)$, where H^s, T^s, and R^s represent the sets of head entities, tail entities, and relations, respectively. The target language fact-relationship graph is denoted as $F^t = (H^t, R^t, T^t)$, where H^t, T^t, and R^t refer to the sets of head entities, tail entities, and relations, respectively. Due to the distinctive linguistic characteristics and syntactic structures of each language, the representation of fact relationships between the two languages is different. Simplifying the process is crucial for tackling the task of creating cross-lingual fact-relationship graphs. This consideration stems from the typical sequential information generation of Transformers when employed as decoders. Therefore, we define a generated graph in the form of a serialized descriptive sentence, which is denoted as $K = (k_1, k_2, \dots, k_n)$. During the training of the model, FGGCLS performs two stages of decoding. In the first stage, a target language fact-relationship graph generator is used to capture both the local semantic features and the global factual relationship features of the source language document and to generate target language factual relationship graphs in the form of a serialized descriptive sentence. In the second stage, a target language summarizer takes the target language fact-relationship graph as a draft and refines it to generate the final target language summary. Therefore, the generating probability distribution of FGGCLS is:

$$P_{(\theta_k, \theta_y)}(Y|D) = P_{\theta_k}(K|D) P_{\theta_y}(Y|D, K) \tag{3}$$

Here, θ_k and θ_y refer to the training parameters for decoding the target language fact-relationship graph and the deliberate decoding of the target summary, respectively (see Fig. 2).

4.2. Encoder

In the present work, our model's encoder consists of a fact-relationship graph encoder and a document encoder. The fact-relationship graph encoder captures global document structure information and semantic representation by modelling the complex fact-relationships within source language documents. Meanwhile, the document encoder encodes the entire document, generates a contextual embedding, and captures local semantic information of the source language document.

4.2.1. Fact-relationship graph encoder

A. Construction of Source Language Fact-Relationship Graph

The fact-relationship graph is a graph network that describes objective logical facts and is composed of fact triples in documents. To construct the source language fact-relationship graph, we first use either the Stanford information extraction tool OpenIE (for English input) or the LTP triplet extraction tool (for Chinese input) to extract fact triples (H^s, R^s, T^s) from the article. The extracted facts are presented as a list of tuples, with each tuple containing a subject H^s, an object T^s, and a relation R^s. Then, we construct a complete relational graph based on these tuples and apply Levi transformation to treat each entity and relation equally. Meanwhile, we include a global node to serve as a connection point for all entity nodes; this enables us to bridge the disconnected parts of the graph and create a more cohesive representation. Specifically, for a given fact triple (H_i^s, R_i^s, T_i^s), we create four nodes H_i^s, T_i^s, R_i^s, and $R_i^s\prime$ and add four directed edges $H_i^s \rightarrow R_i^s$, $R_i^s \rightarrow T_i^s$, $T_i^s \rightarrow R_i^s\prime$, and $R_i^s\prime \rightarrow H_i^s$ to construct the original source language fact-relationship directed graph $F^s = (V, E)$, where V is a list of entities, relations, and a global node, and E is a list of directed edges between the elements in V. During the model training process, we create a source language graph F^s for each CLS training sample, effectively organizing the source language document fact information through the constructed fact-relationship graph.

B. Graph Node Initialization

Consider that a node h is typically composed of a sequence of words $h = \{w_1, w_2, \dots, w_l\}$. To obtain an embedding representation of node

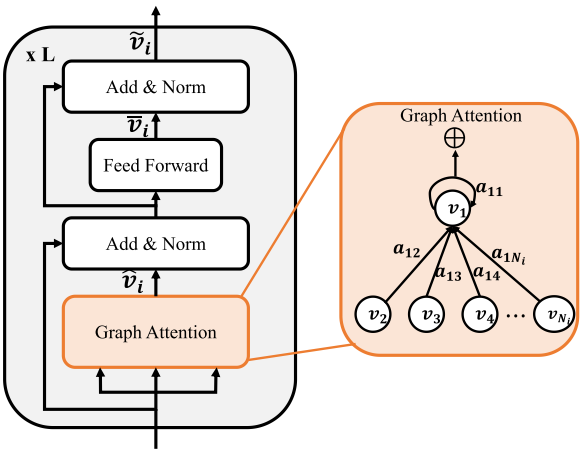

Fig. 3. The Graph Transformer framework is composed of L-layer blocks that comprise Graph Attention, Feed Forward, and LayerNorm.

h, we use positional encoding (PE) to obtain the position information. The final embedding of the ith word is the sum of its word embedding and position embedding. Finally, we input these embeddings into the Self-Attention layer to obtain the initial embedding v_j^0 for each node.

$$v_j^0 = Self\,Attention(\|_{i=1}^{l}(embedding(w_i) + PE(w_i))) \tag{4}$$

Where l is the word sequence length of an entity.

C. Graph Representation Learning

We leverage the Graph Transformer, as proposed by Koncel-Kedziorski et al. [27], to perform hidden representation learning on the nodes of the fact-relationship graph in the source language. The Graph Transformer takes the initialization embedding V^0 as input and updates the representation v_i of each node based on the representations of its neighbouring nodes. To accomplish this, we set the number of self-attention heads in the Graph Transformer to N and calculate N independent heads separately.

$$\hat{v}_i = v_i + \|_{n=1}^{N} \sum_{j \in \varkappa_i} \alpha_{ij}^n W_V^n v_j \tag{5}$$

$$\alpha_{ij}^n = a^n(v_i, v_j) \tag{6}$$

$$a\left(v_i, v_j\right) = \frac{\exp((W_K v_j)^T W_Q v_i)}{\sum_{t \in \aleph_i} \exp((W_K v_t)^T W_Q v_i)} \tag{7}$$

Where $\varkappa_i$ represents the neighbouring nodes in the fact-relationship graph F^s of the source language, while W_Q, W_K, and W_V represent the trainable parameters for the query, key, and value, respectively.

The Graph Transformer is comprised of multi-layer stacked blocks, and each block follows a calculation process as described below:

$$\tilde{v}_i = LayerNorm(\bar{v}_i + LayerNorm(\hat{v}_i)) \tag{8}$$

$$\bar{v}_i = FFN(LayerNorm(\hat{v}_i)) \tag{9}$$

Where FFN is a Feed-Forward neural network.

As shown in Fig. 3, the Graph Transformer is constructed using L layers of stacked blocks. The output from layer $l-1$ is fed as input $v_i^l = \tilde{v}_i^{l-1}$ to layer l. Through multiple iterations of encoding and updating, the final encoding of nodes is represented as $V^L = [v_i^L]$.

4.2.2. Document encoder

While the fact-relationship encoder can capture the global structure and semantic information of a document, it primarily focuses on the factual relationships within the document, which may result in the omission of important local information. To address this issue, we utilize the Transformer's encoder as a document encoder to capture the local semantic information of the source language input document $X = (x_1, x_2, \ldots, x_m)$, thereby reducing the possibility of important local information being overlooked. Specifically, we take the word embedding of the document as input to the document encoder and utilize a multi-head attention block to gather information from the document content at different positions. Each head corresponds to a scaled dot-product attention mechanism:

$$MultiHead = Concat(head_1, head_2, \ldots, head_M)W^O \tag{10}$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{11}$$

$$Attention(QW_i^Q, KW_i^K, VW_i^V) = softmax(\frac{\left(QW_i^Q\right)\left(KW_i^K\right)^T}{\sqrt{d_k}})(VW_i^V) \tag{12}$$

Here, W_i^Q, W_i^K, and W_i^V refer to learnable parameter matrices, and M represents the number of heads for the multi-head attention mechanism.

4.3. Decoder

4.3.1. Fact-relationship graph generator

To model the key information of the text during the decoding process and simplify the complexity of generating cross-language summary sequences, we introduce a target language fact-relationship generator as an extension to the conventional end-to-end sequence generation model; this is the key to transforming the generation problem into a cross-lingual graph generation problem. Specifically, we first construct the target language fact-relationship graph in the form of a serialized descriptive sentence, and then we utilize a target language fact-relationship graph generator to generate this kind of information.

A. Construction of Target Language Fact-Relationship Graph

To create the fact-relationship graph for the target language, we utilize OpenIE or LTP to analyse the target language reference summary, which provides a list of fact triples that can be used as the basis for constructing the fact-relationship graph. To simplify the process of generating the target language fact-relationship graph, we define a standard graph structure in the form of a serialized descriptive sentence that includes entities, entity relationships, and other relevant information. The serialization pattern of the graph structure is determined by the prefix combined with the entity or entity relationship, as illustrated in Fig. 4.

B. Target Language Fact-Relationship Graph Generator

Inspired by the deliberation network, we propose that before decoding the target language summary, it is essential to decode the target language fact-relationship graph based on the source language fact-relationship graph. In other words, the target language fact-relationship graph generator decodes and generates a draft of the target language fact relationship, which is then used by the target language summarizer to generate the final target summary.

The fact-relationship graph encoder captures the structured fact-relationships in the source language document and extracts important global factual features, while the source language transformer encoder captures more local semantic information. To generate the target language fact-relationship graph, the target language fact-relationship graph generator relies on both the graph representation V^L from the source language fact-relationship graph encoder and the source language sequence encoding representation H^L from the document encoder.

The target language fact-relationship graph generator consists of L^t layers of self-attention, an improved cross-attention layer, a feed-forward neural network, and layer normalization. During the decoding

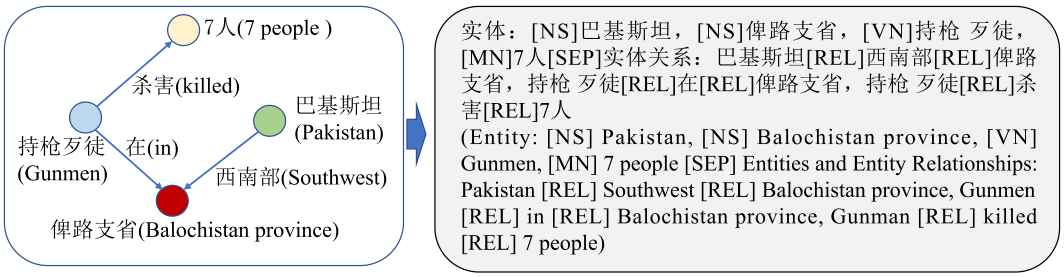

Fig. 4. The fact-relationship graph can be represented as a serialized descriptive sentence using a construction rule that combines a prefix with either an entity or an entity relationship.

process of the target language fact-relationship graph (in the form of a serialized descriptive sentence), the ith self-attention layer encodes the jth token of the target language fact-relationship graph to obtain its representation k_j^i. Then, the jth token representations k_j^i, V^L, and H^L of the graph sequence are used as input for the cross-attention layer to calculate the next token representation. Finally, based on the hidden state representation of the decoder, the prediction of the target language fact-relationship graph is generated through a greedy search.

Within the cross-attention layer, the target language graph sequence representation k_j^i is used as the query Q, while the source language graph representation V^L is used as the key K and value V:

$$c_{j,g}^i = MHAM(k_j^i, V^L, V^L) \tag{13}$$

When using the source language sequence encoding representation H^l as input for the cross-attention layer, the target language graph sequence representation k_j^i is taken as the query Q, while the source language sequence encoding representation H is used as the key K and value V:

$$c_{j,h}^i = MHAM(k_j^i, H^L, H^L)$$

There, MHAM is the multi-head cross attention.

To achieve seamless integration between the source language sequence encoding and the source language graph representation in the target language fact-relationship graph generator, we introduce a target language fusion module based on [30]. The feature fusion mechanism of this module is outlined below:

$$\bar{c}_j^i = Concat(c_{j,g}^i, c_{j,h}^i) \tag{14}$$

$$s_j^i = \tanh\left(W_a^i c_{j,g}^i + W_b^i \bar{c}_j^i\right) \tag{15}$$

$$z_j^i = sigmoid\left(W_c^i c_{j,g}^i + W_d^i \bar{c}_j^i\right) \tag{16}$$

$$c_j^i = \left(1 - z_j^i\right) \odot s_j^i + z_j^i \odot c_{j,g}^i \tag{17}$$

Among them, W_a^i, W_b^i, W_c^i, and W_d^i refer to learnable linear parameters. Then, $\bar{c}_j^i$ will fed into a residual network:

$$\hat{c}_j^i = LayerNorm(c_j^i + FFN(c_j^i)) \tag{18}$$

4.3.2. Target language summarizer

The goal of the target language summarizer is to generate a summary text in the target language by prediction based on the encoder representation and the generation results of the target language fact-relationship graph generator in the first stage. The decoding process is similar to the deliberate process in the deliberation network. The context representation $c_{j,h}^i$ output by the document encoder, the context representation $c_{j,g}^i$ of the source language graph, and the context representation $c_{j,t}^i$ of the result generated by the target language fact-relationship graph generator after greedy decoding are fused as: c_j^i

$$c_j^i = z_1 * \hat{c}_{j,h}^i + \left(1 - z_1\right) * \hat{c}_{j,g}^i \tag{19}$$

$$z_1 = sigmoid\left(\left[\hat{c}_{j,h}^i, \hat{c}_{j,g}^i\right] W_{f1} + b_{f1}\right) \tag{20}$$

$$\hat{c}_j^i = z_2 * c_j^i + \left(1 - z_2\right) * c_{j,t}^i \tag{21}$$

$$z_2 = sigmoid\left(\left[c_j^i, c_{j,t}^i\right] W_{f2} + b_{f2}\right) \tag{22}$$

5. Experiments

5.1. Datasets

We evaluate our method on the En2ZhSum and Zh2EnSum CLS datasets released by Zhu et al. [1]. En2ZhSum is an English–Chinese cross-lingual dataset containing 364,687 English documents (average of 755 tokens) and a Chinese summary (average of 55 tokens). The dataset is divided into 364,687 training samples, 3000 validation samples, and 3000 test samples. This dataset is constructed from CNN/DM [31] and MSMO [32] using a round-trip translation strategy. Zh2EnSum is a Chinese–English summarization dataset containing 1,699,713 short Chinese texts (average of 104 tokens) and English short summaries (average of 14 tokens). The dataset is divided into 1,693,713 training samples, 3000 validation samples, and 3000 testing samples. Zh2EnSum is translated from the LCSTS dataset. All training examples contain a source language document and a target language summary.

5.2. Experimental setup

We follow the vocabulary size and text length truncation settings of Zhu et al. [1]. We convert all English characters to lowercase. We set the input source language document truncation length to 200, the Chinese output truncation length to 150, and the English truncated length of the output to 120. We initialize all parameters through the Xavier initialization method. All encoders and decoders have 6 layers and the hidden representation has 512 dimensions. During training, we use the Adam optimizer [33] with $\beta_1 = 0.9$, $\beta_2 = 0.998$, and $lr = 10^{-9}$. We train our model using two NVIDIA 3090 GPUs and reach convergence in 1 million iterations. Our fact-relationship graph generator is greedily searched during testing, and the summary decoder is generated using beam search with beam size 4.

5.3. Baselines

We compare the proposed FGGCLS model with two pipeline-based methods and current advanced end-to-end models:

GETran: A pipeline method based on Google Translator, first translating and then summarizing. GETran first translates the original document into the target language by Google Translator. Then, GETran summarizes the translated document using LexRank [34], a powerful and widely used unsupervised summarization method.

GLTran: A pipeline method based on Google Translator that summarizes before translation. First, a summarization model is trained

Table 1
The F1 score (%) of ROUGE-1, ROUGE-2 and ROUGE-L for En2ZhSum and Zh2EnSum datasets.

Method	En2ZhSum			Zh2EnSum		
	RG-1	RG-2	RG-L	RG-1	RG-2	RG-L
GETran	28.24	11.27	25.49	24.31	8.77	19.84
GLTran	31.75	13.50	28.32	33.47	16.24	28.93
TNCLS	35.21	16.79	31.70	38.35	21.28	33.51
CLS+MS	36.72	18.16	33.17	39.69	22.55	34.80
ATS-N	37.18	18.52	33.58	39.62	22.63	34.77
ATS-E	37.05	18.36	33.45	39.38	22.45	34.08
ATS-A	37.49	18.83	33.89	39.87	22.92	34.96
FGGCLS	**38.71**	**19.82**	**34.93**	**40.51**	**23.59**	**35.43**

Table 2
Ablation study. "FGGCLS w/o FE" refers to the version of the model that lacks the fact-relationship graph encoder, while "FGGCLS w/o FG" refers to the variant that omits the target language fact-relationship graph generator.

Method	En2ZhSum			Zh2EnSum		
	RG-1	RG-2	RG-L	RG-1	RG-2	RG-L
TNCLS	35.21	16.79	31.70	38.35	21.28	33.51
FGGCLS w/o FE	36.87	18.07	32.75	39.93	22.82	34.63
FGGCLS w/o FG	37.96	18.83	33.69	40.24	23.15	35.09
FGGCLS	**38.91**	**20.02**	**34.73**	**40.51**	**23.59**	**35.43**

based on the source language monolingual summarization dataset. The source language documents are fed into the monolingual summarization model to generate source language summaries. Based on this, the summaries are translated into the target language by Google Translate.

TNCLS: The first end-to-end cross-lingual summary model was proposed by Zhu et al. [1] To reduce the complexity of the model, they initialized the model by a random initialization algorithm.

CLS+MS: A multi-task learning model based on NCLS. The MS model is added to the traditional end-to-end framework to train together.

ATS-N: Zhu et al. [35] proposes an end-to-end CLS model based on a heuristic translation model. The approach decomposes cross-lingual summarization into three steps: attend, translate, and summarize. For the "translate" step, the authors compare three strategies: Naive, Equal, and Adapt. We define ATS-E as using the naive strategy.

ATS-E: Similar to the ATS-N, except that we replace the naive strategy in the "Translate" step with the Equal strategy.

ATS-A: Similar to the ATS-N, except that the naive strategy is replaced by the adaptive strategy in the "translate" step.

Recently, some other models have shown improvements in the performance of CLS tasks but require additional training data or pre-trained models in advance, and therefore cannot be fairly compared with FGGCLS. For example, the MCLAS method proposed by Bai et al. [4] requires an additional source language summary as reference summary and pre-trained the MCLAS model in advance based on a large-scale monolingual summary dataset, which increases the complexity of model training; the CLS+MT proposed by Zhu et al. [1] training method relies on additional large-scale MT datasets; and the VHM method proposed by Liang et al. [5] also requires large-scale summary and machine translation datasets. Therefore, we do not use these models as baselines.

5.4. Experimental results and analysis

We evaluate all models using the standard ROUGE metric on all datasets, reporting F1 scores for ROUGE-1, ROUGE-2, and ROUGE-L. All ROUGE scores are reported by the 95% confidence interval of the official script measure.

FGGCLS vs. baselines. We reimplement the GETran, GLTran, TNCLS, CLS+MS and ATS models on En2ZhSum and Zh2EnSum using word–word segmentation granularity. The experimental results are

Table 3
Human Evaluation of faithfulness using a 3-star rating system. The system rates major factual error as 1, minor factual error as 2, and no factual error as 3. A rating of 3 is the most faithful, and a higher percentage of ratings of 3 indicates better faithfulness.

Model	En2ZhSum			Zh2EnSum		
	1	2	3	1	2	3
TNCLS	42.00	27.00	31.00	41.50	14.00	44.50
ATS-A	34.00	28.00	38.00	38.00	13.00	49.00
FGGCLS	**31.50**	**22.00**	**46.50**	**34.50**	**14.50**	**51.00**

Table 4
Human evaluation results, with IF representing informativeness, CC representing conciseness, and FL representing fluency.

Model	En2ZhSum			Zh2EnSum		
	IF	CC	FL	IF	CC	FL
TNCLS	3.04	3.24	3.12	3.36	3.80	3.76
ATS-A	3.20	**3.64**	3.48	3.88	**4.04**	4.08
FGGCLS	**3.48**	3.52	**3.56**	**4.12**	3.96	**4.16**

shown in Table 1. As shown in Table 1, we can observe that FGGCLS significantly outperforms the baselines on En2ZhSum. It is worth noting that FGGCLS has a more significant performance improvement relative to the first end-to-end CLS model TNCLS, reaching +3.50 ROUGE-1, +3.03 ROUGE-2, and +3.23 ROUGE-L. Compared with CLS+MS, ATS-A and other advanced CLS methods, FGGCLS does not use additional monolingual summary datasets or probabilistic bilingual dictionaries, significantly reducing the data dependence.

FGGCLS also performs better than the baselines on the Zh2EnSum. However, the improvement of FGGCLS in Zh2EnSum is less noticeable compared to the En2ZhSum dataset; this is because the input document length of the Zh2EnSum is shorter and contains less information about the fact triples, resulting in a sparse fact-relationship graph. The short-text CLS dataset is noisier. The extraction performance of fact triples by OpenIE is limited, which also limits the performance of FGGCLS.

Overall, the experimental results on En2ZhSum and Zh2EnSum show that our proposed FGGCLS significantly outperforms the baseline on longer-length news texts and achieves satisfactory performance on short textual CLS datasets. In addition, FGGCLS can relax the dependence of the model on the dataset and be easily generalized to more CLS tasks.

Ablation study. To demonstrate the impact of our proposed fact-relationship encoder (FE) and fact-relationship generator (FG) on model performance. We use two datasets, En2ZhSum and Zh2EnSum, to train our models and compare their ROUGE values. The results are shown in Table 2. On both datasets, FGGCLS, FGGCLS w/o FE, and FGGCLS w/o FG significantly outperform the baseline model NCLS, indicating that the proposed cross-lingual fact-relationship graph encoder and generation method has a significant positive impact on the CLS task. Meanwhile, we find that FGGCLS w/o FE performs much better than FGGCLS w/o FG, which indicates that the target language fact-relationship generator contributes more to the model.

In addition, there are some differences in the performance of our proposed modules on different datasets. On the En2ZhSum, the performance degradation of the model FGGCLS w/o FE with the fact relation encoder removed is more significant, reaching −2.04 ROUGE-1. In contrast, on ZH2ENsum, the performance degradation of the model FGGCLS w/o FE with the fact relation encoder removed is not as that of En2ZhSum, which may be caused by the input document length of En2ZhSum being longer and having more redundant information, while the fact-relationship encoder can effectively filter invalid information in the encoding stage and strengthen the encoder's ability to capture critical fact-relationships.

Input : 7 月 7 日，上海证券交易所和深圳证券交易所分别发布《上海证券交易所股票上市规则》、《深圳证券交易所股票上市规则》，对原规则中的退市、停复牌等内容进行了修订。上述两个新规则自发布之日起施行。 (Translation: **On July 7**, **the Shanghai Stock Exchange and the Shenzhen Stock Exchange** issued the "Shanghai Stock Exchange Stock Listing Rules" and "Shenzhen Stock Exchange Stock Listing Rules," revising the original rules on **delisting**, suspension, and resumption of trading, etc... The above **two new rules** will be enacted on the promulgation date.)
Translation Summary: three points of attention in implementing new listing regulations on **shanghai and shenzhen stock exchanges**.
Gold Summary: the three highlights of the **shanghai and shenzhen stock exchange's** implementation of the **new listing regulations** are worthy of attention.
TNCLS: **the shanghai stock exchange** promulgates rules for the listing of **the new stock exchange**: suspension of stock transfer system and clearing shares in collective stock market.
ATS-A: the **new rules** of **the shanghai stock exchange** and the implementation of the new listing rules on the shanghai stock exchange have been implemented.
FGGCLS (serialization fact-relationship graph): Entity: [SUB] **shanghai and shenzhen**, [OBJ] **two new rules**, [SUB] **new delisting**, [OBJ] **revised** [SEP] Entities and Entity Relationships: **shanghai and shenzhen** [REL] **promulgated** [REL] **two new rules**, **new delisting** [REL] have [REL] **revised**
FGGCLS: two new rules of **listing issued** by **shanghai and shenzhen stock exchange** have been promulgated and implemented since **july 7**, and the new **delisting rules** have been **revised**.

Fig. 5. An example of a cross-language summary generated by different models, with significant entities and relations highlighted in bold.

Table 5
Graph generation quantitative analysis results. Human evaluation of the knowledge existence and faithfulness of the results generated by the fact-relationship graph generator.

Method	En2ZhSum		Zh2EnSum	
	Knowledge Existence	Faithfulness	Knowledge Existence	Faithfulness
FGGCLS	**3.31**	**3.08**	**3.84**	**3.16**
FGGCLS w/o FE	2.92	2.80	3.30	2.95

5.5. Human evaluation

5.5.1. Faithfulness

In addition to automatically evaluating the CLS model, we also conduct a human evaluation to verify the faithfulness of the summaries generated FGGCLS. We randomly select 100 samples from the test datasets of En2ZhSum and Zh2EnSum. Five graduate students with excellent English and Chinese literacy skills are recruited to independently rate the faithfulness of all 100 summary samples generated by the TNCLS, ATS-A, and FGGCLS models. For the faithfulness evaluation criteria, we ask them to follow a 3-star rating (1 = major factual error, 2 = minor factual error, 3 = no factual error). A majority vote is then used to aggregate the three judgments for each summary. We show the distribution of summaries with ratings of 1, 2, and 3 stars in Table 3.

We can observe that the target language summaries generated by FGGCLS are more reliable than those generated by TNCLS and ATS-A, and FGGCLS's target language summaries are closer to the facts themselves. Specifically, FGGCLS achieves 46.50% and 51% no factual error, surpassing TNCLS by 6.5% and 5.5%. We also observe that FGGCLS can significantly reduce major factual errors, especially on the Zh2EnSum dataset, by 3.5% for FGGCLS relative to the ATS-A model, effectively improving the factual consistency of the summaries generated by the CLS task.

5.5.2. Informativeness, fluency and conciseness

In addition to the faithfulness evaluation, we randomly select 25 samples from the En2ZhSum and Zh2EnSum test sets and compare the summaries generated by the TNCLS, ATS-A, and FGGCLS models. We recruit three graduate students to evaluate informativeness (IF), fluency (FL) and conciseness (CC), each scored from 1 (worst) to 5 (best). The results are shown in Table 4.

The informativeness score, conciseness score and fluency score of FGGCLS are significantly better than those of the baseline model TNCLS, which further proves the effectiveness of our proposed method. The conciseness score of FGGCLS is comparable to that of ATS-A, but the summary text generated by FGGCLS is more informative and has higher text fluency.

5.5.3. Graph generation quantitative analysis

To verify the role of the fact-relationship graph in our model, we conduct a qualitative evaluation of knowledge existence and faithfulness in the results generated by the fact-relationship graph generator.

Input : 2 日上午，一名华人官员在菲北部卡加延省土格加劳市遭枪击身亡。当晚，两名 20 多岁的华商遭不明身份枪手杀害。我使馆已联系菲警方，要求尽快核实遇害者身份信息。再次提醒：中国公民近期暂勿前往菲律宾！在菲中国公民注意安全！ (Translation: On the morning of the 2nd, a Chinese official was shot and killed in Tuguegarao City, located in Cagayan Province, northern Philippines. On the night, a unidentified gunmen killed two 20-year-old Chinese businessmen. Our embassy has contacted the Philippine police, urging them to expedite the verification of the victims' identities. Again: Chinese citizens are advised to avoid traveling to the Philippines in the near future ! Chinese citizens currently in the Philippines are reminded to pay attention to their safety !)
Translation Summary: three chinese in the philippines were shot in a day.
Gold Summary: three chinese in the philippines were shot in a day.
TNCLS: three chinese in the philippines were shot dead in a shooting incident near the philippine town , huangyan island, claiming that there was no chinese .
ATS-A: three chinese in the philippines were shot in a day and three died in a shooting in the country.
FGGCLS (serialization fact-relationship graph): Entity: [SUB] two, [OBJ] middle [SEP] Entities and Entity Relationships: two [REL] were in [REL] middle
FGGCLS: three chinese in the philippines were shot in a day, two of whom were in the middle of the night !

Fig. A.6. Example-1 of a cross-language summary generated by different models on Zh2En dataset.

Input : 俄罗斯央行 22 日宣布，将对特拉斯特国有银行进行财政整顿，并指定俄罗斯银行存款保险局对该银行业务进行临时托管，以避免该银行破产。这是俄罗斯在本轮卢布危机中救助的第一家银行。 (Translation: On the 22nd, the Central Bank of Russia announced that it will carry out financial consolidation of the state-owned bank of Trast, and designate the Bank of Russia Deposit Insurance Agency to temporarily manage the banking business to avoid the bank's bankruptcy. This is the first bank rescued by Russia in the current ruble crisis.)
Translation Summary: Russia announces the rescue of a state-owned bank .
Gold Summary: russia announces the rescue of a state-owned bank .
TNCLS: russian media : the russian central bank proposes to bankruptcy the liquidation mechanism of bank assets has been carried out by the banking regulatory commission .
ATS-A: the russian central bank announced the closure of the bank financial deposit insurance bureau for the first time in 25 years , and the bank of russia took over .
FGGCLS (serialization fact-relationship graph): Entity: [SUB] central bank, [OBJ] banking industry, [SUB] banking industry, [OBJ] crisis [SEP] Entities and Entity Relationships: central bank [REL] intends to [REL] banking industry, banking industry [REL] save bank in [REL] crisis
FGGCLS: the russian central bank intends to rectify the banking industry and save the first bank in the crisis of bankruptcy in the rouble crisis .

Fig. A.7. Example-2 of a cross-language summary generated by different models on Zh2En dataset.

Input : time is running out for 18 miners trapped underground after a coal mine collapsed in southern turkey on tuesday. governor murat koca said the incident occurred at the has sekerler mine near the town of ermenek in karaman province , 300 miles south of ankara close to turkey's mediterranean coast . . a rescue team has been working all night and into wednesday to free the trapped miners , who are believed to be located about 350 meters below ground . . aid locals watch as rescue workers try to save 18 miners trapped after a collapse in a mine in the ermenek district in karaman , turkey . turkish officials say the workers' chances of survival are slim unless they managed to reach a safety gallery . the flood occurred around 3pm local time when workers were eating their lunch , kerim pinarli , one of 16 miners who was able to escape , told ntv news channel . 'we smelled gas and heard our friends below shout , there's gas , don't come ! we escaped by seconds , ' pinarli said . mr koca said the miners were trapped after water accumulated inside ...
Translation Summary: 土耳其一座煤矿坍塌，18 名工人被困在地下 300 米。 (Translation: A coal mine in Turkey has collapsed, trapping 18 workers 300m underground .)
Gold Summary: 土耳其一座煤矿坍塌，18 名工人被困在地下 300 米。 (Translation: A coal mine in Turkey has collapsed, trapping 18 workers 300m underground .)
TNCLS: 土耳其中部城镇班尼斯的煤矿发生塌方，导致 300 多名工人被困在洪水中。 (Translation: In the central Turkish town of Bannis, a coal mine has collapsed, trapping over 300 workers in floodwaters .)
ATS-A: 18 名工人因内部积水而被困在土耳其一个矿井里。 (Translation: Due to flooding inside, 18 workers have been trapped in a mine in Turkey .)
FGGCLS (serialization fact-relationship graph): 实体：[N]一 处 煤矿，[V]塌方，[N]18 名 矿工，[ND]土耳其 东南部 [SEP] 实体关系：一 处 煤矿[REL]发生[REL]塌方，18 名 矿工[REL]困在[REL]土耳其 东南部 (Translation: Entity: [N] a coal mine, [V] collapse, [N] 18 miners, [ND] southeastern Turkey [SEP] Entities and Entity Relationships: a coal mine [REL] a [REL] collapse, 18 miners [REL] trapped in [REL] southeastern Turkey)
FGGCLS: 土耳其当局称，一处煤矿发生塌方，18 名矿工被困在土耳其东南部的一个矿井里。 (Translation: Turkish authorities said that 18 miners have been trapped in a mine shaft in southeastern Turkey after a coal mine collapsed .)

Fig. A.8. Example-3 of a cross-language summary generated by different models on En2Zh dataset.

A subset of 50 samples is randomly selected from the En2ZhSum and Zh2EnSum test sets. Three graduate students are recruited to perform a qualitative assessment of the fact-relationship graph generated by the FGGCLS and FGGCLS w/o FE methods using a scoring system ranging from 1 (worst) to 5 (best). The evaluation results are presented in Table 5.

We have observed that FGGCLS performs better than FGGCLS w/o FE in terms of knowledge existence and faithfulness. This observation indicates that the fact-relationship graph generator can generate more practical knowledge while ensuring factual consistency.

5.6. Case study

We construct a case study of the Zh2EnSum test set. The target language summaries generated by each model are shown in Fig. 5 (more case studies can be found in the Appendix).

Comparison with the manually labelled Gold Summary, the TNCLS generates incorrect information such as “the shanghai stock exchange” and “suspension of stock transfer system”, leading to serious factual errors. The ATS-A model loses “shenzhen stock exchange” in describing the stock exchange and suffers from fluency and repeatability problems.

In contrast, the serialization fact-relationship graph that is generated by FGGCLS contains several key entities and entity relationships such as "[SUB] shanghai and shenzhen" and "new delisting [REL] have [REL] revised", which help FGGCLS generate a summary that covers almost all points and generates key events and time information, such as "shanghai and shenzhen" and "delisting rules have been revised ". In conclusion, FGGCLS can generate more accurate English summaries than the baselines.

6. Conclusion and discussion

In this paper, we present an innovative approach to address the cross-lingual summarization (CLS) task, termed Cross-Lingual Fact Graph Generation for Cross-Lingual Summarization (FGGCLS). This approach transforms the conventional CLS problem into a structured graph generation task. By explicitly modelling the intricate fact information present in the source language documents and the target language reference summaries, we convert the source language facts into target language facts using a cross-lingual fact graph generation method. Therefore, the challenges of semantic alignment and information compression in the CLS model. Furthermore, we leverage the fact graph of the generated target language summary as a draft, enabling our model to generate a more accurate and reliable target language summary. Our experiments on the Zh2EnSum and En2ZhSum datasets demonstrate that FGGCLS effectively enhances the performance of CLS and reduces factual inconsistencies in cross-lingual summarization generation. This shows that complex relational graphs are an effective way to improve the performance of CLS.

Furthermore, there are various promising avenues for further research based on the findings of this study. For example, we can extend our cross-lingual fact-relationship graph generation approach to other cross-lingual generation tasks (e.g., cross-lingual dialog summarization and cross-lingual multi-document summarization) and to explore the impact of different relationship types on the model.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data

Appendix. More case studies

To provide more qualitatively support for our claim, we present three more results of case studies as show in Figs. A.6–A.8.

References

[1] J. Zhu, Q. Wang, Y. Wang, Y. Zhou, J. Zhang, S. Wang, C. Zong, NCLS: Neural cross-lingual summarization, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3045–3055, http://dx.doi.org/10.18653/v1/D19-1302, URL: https://www.aclweb.org/anthology/D19-1302.

[2] S. Takase, N. Okazaki, Multi-task learning for cross-lingual abstractive summarization, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2022, pp. 3008–3016.

[3] Y. Cao, H. Liu, X. Wan, Jointly learning to align and summarize for neural cross-lingual summarization, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 6220–6231.

[4] Y. Bai, Y. Gao, H.-Y. Huang, Cross-lingual abstractive summarization with limited parallel resources, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 6910–6924.

[5] Y. Liang, F. Meng, C. Zhou, J. Xu, Y. Chen, J. Su, J. Zhou, A variational hierarchical model for neural cross-lingual summarization, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022, pp. 2088–2099.

[6] S.-q. Shen, Y. Chen, C. Yang, Z.-y. Liu, M.-s. Sun, et al., Zero-shot cross-lingual neural headline generation, IEEE/ACM Trans. Audio Speech Lang. Process 26 (12) (2018) 2319–2327.

[7] T.T. Nguyen, A.T. Luu, Improving neural cross-lingual abstractive summarization via employing optimal transport distance for knowledge distillation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, no. 10, 2022, pp. 11103–11111.

[8] F. Boudin, S. Huet, J.-M. Torres-Moreno, A graph-based approach to cross-language multi-document summarization, Polibits (43) (2011) 113–118.

[9] X. Wan, Using bilingual information for cross-language document summarization, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, pp. 1546–1555.

[10] J. Zhang, Y. Zhou, C. Zong, Abstractive cross-language summarization via translation model enhanced predicate argument structure fusing, IEEE/ACM Trans. Audio Speech Lang. Process. 24 (10) (2016) 1842–1853.

[11] E. Linhares Pontes, S. Huet, J.-M. Torres-Moreno, A.C. Linhares, Cross-language text summarization using sentence and multi-sentence compression, in: Natural Language Processing and Information Systems: 23rd International Conference on Applications of Natural Language To Information Systems, NLDB 2018, Paris, France, June 13-15, 2018, Proceedings 23, Springer, 2018, pp. 467–479.

[12] X. Wan, F. Luo, X. Sun, S. Huang, J.-g. Yao, Cross-language document summarization via extraction and ranking of multiple summaries, Knowl. Inf. Syst. 58 (2019) 481–499.

[13] J. Ouyang, B. Song, K. McKeown, A robust abstractive system for cross-lingual summarization, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 2025–2031, http://dx.doi.org/10.18653/v1/N19-1204, URL: https://aclanthology.org/N19-1204.

[14] C. Orăsan, O.A. Chiorean, Evaluation of a cross-lingual romanian-english multi-document summariser, 2008.

[15] X. Wan, H. Li, J. Xiao, Cross-language document summarization based on machine translation quality prediction, in: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010, pp. 917–926.

[16] F. Ladhak, E. Durmus, C. Cardie, K. Mckeown, WikiLingua: A new benchmark dataset for cross-lingual abstractive summarization, in: Findings of the Association for Computational Linguistics: EMNLP 2020, 2020, pp. 4034–4048.

[17] A. Leuski, C.-Y. Lin, L. Zhou, U. Germann, F.J. Och, E. Hovy, Cross-lingual c* st* rd: English access to hindi information, ACM Trans. Asian Lang. Inf. Process. (TALIP) 2 (3) (2003) 245–269.

[18] J.-g. Yao, X. Wan, J. Xiao, Phrase-based compressive cross-language summarization, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 118–127.

[19] X. Duan, M. Yin, M. Zhang, B. Chen, W. Luo, Zero-shot cross-lingual abstractive sentence summarization through teaching generation and attention, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 3162–3172.

[20] D. Jin, C. Huo, C. Liang, L. Yang, Heterogeneous graph neural network via attribute completion, in: Proceedings of the Web Conference 2021, 2021, pp. 391–400.

[21] D. Jin, Z. Yu, P. Jiao, S. Pan, D. He, J. Wu, P. Yu, W. Zhang, A survey of community detection approaches: From statistical modeling to deep learning, IEEE Trans. Knowl. Data Eng. (2021).

[22] M. Chen, W. Li, J. Liu, X. Xiao, H. Wu, H. Wang, SgSum: Transforming multi-document summarization into sub-graph selection, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021, pp. 4063–4074.

[23] P. Cao, J. Wu, GraphRevisedIE: Multimodal information extraction with graph-revised network, Pattern Recognit. (2023) 109542.

[24] Z. Yu, D. Jin, Z. Liu, D. He, X. Wang, H. Tong, J. Han, AS-GCN: Adaptive semantic architecture of graph convolutional networks for text-rich networks, in: 2021 IEEE International Conference on Data Mining, ICDM, IEEE, 2021, pp. 837–846.

[25] P. Fernandes, M. Allamanis, M. Brockschmidt, Structured neural summarization, in: International Conference on Learning Representations.

[26] A. Fan, C. Gardent, C. Braud, A. Bordes, Using local knowledge graph construction to scale Seq2Seq models to multi-document inputs, in: 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, 2019.

[27] R. Koncel-Kedziorski, D. Bekal, Y. Luan, M. Lapata, H. Hajishirzi, Text generation from knowledge graphs with graph transformers, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 2284–2293.
[28] S. Jiang, D. Tu, X. Chen, R. Tang, W. Wang, H. Wang, CptGraphSum: Let key clues guide the cross-lingual abstractive summarization, 2022, arXiv preprint arXiv:2203.02797.
[29] Y. Xia, F. Tian, L. Wu, J. Lin, T. Qin, N. Yu, T.-Y. Liu, Deliberation networks: Sequence generation beyond one-pass decoding, Adv. Neural Inf. Process. Syst. 30 (2017).
[30] X. Chen, H. Alamro, M. Li, S. Gao, X. Zhang, D. Zhao, R. Yan, Capturing Relations Between Scientific Papers: an Abstractive Model for Related Work Section Generation, Association for Computational Linguistics, 2021.
[31] R. Nallapati, B. Zhou, C. dos Santos, Ç. Gülçehre, B. Xiang, Abstractive text summarization using sequence-to-sequence RNNs and beyond, in: Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, 2016, pp. 280–290.
[32] J. Zhu, H. Li, T. Liu, Y. Zhou, J. Zhang, C. Zong, MSMO: Multimodal summarization with multimodal output, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 4154–4164.
[33] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Y. Bengio, Y. LeCun (Eds.), ICLR (Poster), 2015, URL: http://dblp.uni-trier.de/db/conf/iclr/iclr2015.html#KingmaB14.
[34] G. Erkan, D.R. Radev, Lexrank: Graph-based lexical centrality as salience in text summarization, Journal of artificial intelligence research 22 (2004) 457–479.
[35] J. Zhu, Y. Zhou, J. Zhang, C. Zong, Attend, translate and summarize: An efficient method for neural cross-lingual summarization, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 1309–1321, URL: https://www.aclweb.org/anthology/2020.acl-main.121.

Yongbing Zhang received the M.S. degree in Network Engineering from Yunnan Normal University. Currently, he is currently pursuing a Ph.D. degrees in Computer Science and Technology at Kunming University of Science and Technology. His research interests include machine learning, nature language processing and information retrieval.

Shengxiang Gao received the M.S. degree in Pattern Recognition and Intelligent System and the Ph.D. degree in Control Engineering from Kunming University of Science and Technology in 2005 and 2016, respectively. She is currently associate professor in School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China. Her main research interests include machine learning, nature language processing and machine translation.

Yuxin Huang received the Ph.D. degree from Kunming University of Science and Technology in 2021. Now he is an associate professor at Kunming University of Science and Technology. His research interests include natural language processing, text generation etc.

Kaiwen Tan received the Ph.D. degree from the Communication and Computer Network Lab of Guangdong, School of Computer Science and Engineering, South China University of Technology in 2021. Currently, he is an instructor in the School of Information Engineering and Automation, Kunming University of Science and Technology, China. His main research interests are natural language processing and bioinformatics.

Zhengtao Yu received his Ph.D. degree in computer application technology from Beijing Institute of Technology, Beijing, China, in 2005. He is currently a professor in the School of Information Engineering and Automation, Kunming University of Science and Technology, China. His main research interests include natural language processing, information retrieval and machine learning.