A Text Summarization Model Based on Dual Pointer Network Fused with Keywords

Ning Ouyang, Xingke Du and Xiaodong Cai
School of Information and Communication, Guilin University of Electronic Technology, Guangxi, China

Keywords: Dual Pointer, Keywords, MTT.

Abstract: In text summarization tasks, the importance of keywords in the text is often overlooked, resulting in the generated summary deviating from the original meaning of the text. To address this issue, a text summarization generation model GMDPK (Generation Model based on Dual Pointer network fused with Keywords) is proposed. Firstly, we design a keyword extraction module MTT (Module based on Topic Awareness and Title Orientation). It enriches semantic features by mining potential themes, and uses highly summarized and valuable information in the title to guide keyword generation, resulting in a summary that is closer to the original meaning of the text. In addition, we add an ERNIE pretraining language model in the word embedding layer to enhance the representation of Chinese text syntax structure and entity phrases. Finally, a keyword information pointer is added to the original single pointer generation network, forming a dual pointer network. This helps to improve the coverage of the copying mechanism and more effectively mine keyword information. Experiments were conducted on the Chinese dataset LCSTS, and the results showed that compared to other existing models, the summary generated by GMDPK can contain more key information, with higher accuracy and better readability.

1 INTRODUCTION

Text summarization refers to extracting key information from the source text, converting it into a brief summary containing key information. It can compress cumbersome text, extract key information and thematic content from the original text, and quickly extract the required information from the text.

Automatic text summarization is divided into extractive and generative summarization methods. The extractive summarization method determines the importance of each sentence in the original text, extracts the top few sentences with higher importance, and reorders them to form a summary. The generative summarization method is to generate creative summaries based on understanding the original text by analyzing its grammar and semantics. With the rapid development of deep learning, (Sutskever et al., 2014) successfully applied the seq2seq model to NLP tasks to solve translation problems in 2014. Inspired by this, (Rush et al., 2015) introduced the seq2seq model into generative summarization tasks, proposed using a convolutional model with attention mechanism to encode documents, and then used a feedforward network-based neural network language model as a decoder to generate summaries. Compared with traditional methods, this model achieved significant performance improvement on the DUC-2004 (Over et al., 2007) and GigaWord datasets. Subsequently, (Napoles et al., 2012) made several improvements to the seq2seq based on RNN, improving the quality of generated summaries. The above methods indicate that seq2seq performs well in generative text summarization, but the generated summaries have many issues with OOV and poor readability. To address the above issues, (Nallapati et al., 2016) proposed the CopyNet model, which allows for the copying of subsequences when the OOV appears. This model effectively reduces the number of OOV and improves the quality of summaries on the Chinese dataset LCSTS (Hu et al., 2015). Subsequently, (See et al., 2017) improved the copy mechanism and proposed a pointer generation model, which uses pointers to choose copying original text words or generating vocabulary words. This model reduces the problem of duplicate summaries.

Du, X., Ouyang, N. and Cai, X.
A Text Summarization Model Based on Dual Pointer Network Fused with Keywords.
DOI: 10.5220/0012881900004536

In *Proceedings of the 1st International Conference on Data Mining, E-Learning, and Information Systems (DMEIS 2024)*, pages 82-86
ISBN: 978-989-758-715-3

The summarization method based on seq2seq may result in the loss of key information from the original text in the generated summary. In order to highlight the importance of keyword information in abstract generation, many studies use keyword information to improve the quality of model generated summaries. (Wan et al., 2007) improved the quality of abstracts by simultaneously extracting abstracts and keywords from a single document, assuming that they can mutually enhance each other. In recent years, (Xu et al., 2020) proposed a guided generation model that combines extraction and abstraction methods, using keyword calculation attention distribution to guide summary generation. (Li et al., 2020) adopted a keyword guided selective encoding strategy to filter source information by studying the interaction between input sentences and keywords, and achieved good results on English datasets The above methods indicate that keywords are beneficial for generating text summaries in English datasets.

In this article, in order to highlight the importance of keywords, we propose a text summarization generation model GMDPK.Firstly, the ERNIE pretraining language model is used to obtain the vector of multi-dimensional semantic features of the article. Then, a keyword generation module MTT is designed to extract keywords. Then, the extracted keyword information is used to calculate attention with the original text. This attention is added to the attention mechanism of the pointer generation network to form a dual pointer network, making the generated abstract pay more attention to the key information of the original text, Improve the accuracy and readability of the abstract.

2 KEYWORD EXTRACTION MODULE BASED ON TOPIC AWARENESS AND TITLE ORIENTATION

For keyword extraction tasks, traditional methods of word frequency statistics cannot effectively reflect the information of keywords. Our MTT is based on the fusion of topic perception and title orientation, which includes a topic extraction neural module, a title oriented hierarchical encoder, and a keyword decoder. The module takes the original text as input and keywords as output. The MTT is shown in Figure 1.

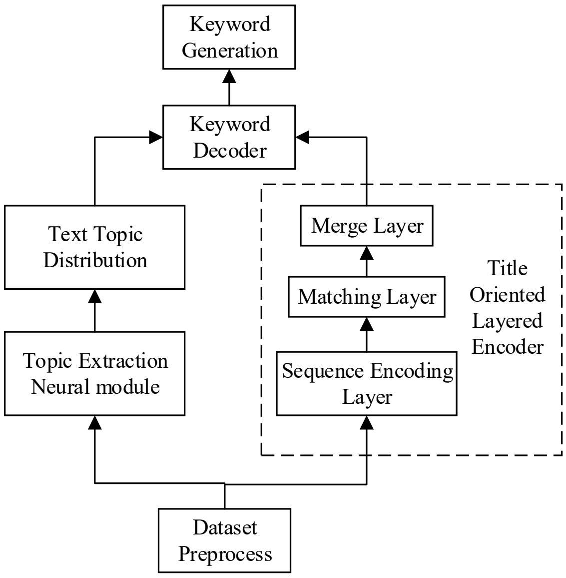

Figure 1: Keyword extraction module based on topic awareness and title orientation.

2.1 Topic Extraction Neural Module

Process each text word in the article into a word vector X_b and input it into the topic extraction neural module. The BoW (Akuma et al., 2022) encoder in the module estimates the prior variables μ 、σ of each word vector, and the formula is as follows:

$$\mu = \int_{\mu}(\int_{e}(X_{b})) \tag{1}$$

$$\log\sigma = \int_{\sigma}(\int_{e}(X_{b})) \tag{2}$$

Where $f(\cdot)$ is a neural perceptron with RuLU activation function, Through the BoW decoder, prior variables are used for topic representation. $Z \sim N(\mu, \sigma^2)$. Constructing topic mixing vector θ Using Softmax Function to guide keyword generation.

$$\theta = softmax(w_\theta Z) \tag{3}$$

2.2 Title Oriented Layered Encoder

As shown in Figure 1, the title oriented hierarchical encoder consists of a sequence encoding layer, a matching layer, and a merging layer.

The sequence encoding layer reads the title input and body content input, and learns their contextual representations through two bidirectional gate recursive units (GRUs).

The matching layer with attention is used to aggregate relevant information from each word title in the context to form an aggregated information vector for each word, and the matching layer is also composed of two parts. One part is the self matching from title to title, and the other part is the matching from text to title.

The merge layer inputs the context vector and aggregate information vector into the merge layer to obtain a title oriented context representation M.

2.3 Keyword Decoder

Input the topic mixing vector θ and title context representation M into the decoder to obtain the probability of keyword Y_i generation.

$$P(\mathrm{Y_i}) = P(\mathrm{M}/\theta) \quad (4)$$

3 A TEXT SUMMARIZATION MODEL BASED ON DUAL POINTER NETWORK FUSION OF KEYWORDS

In this section, we will introduce the text summarization model based on dual pointer network fusion of keywords (GMDPK). Figure 2 is the network model diagram.

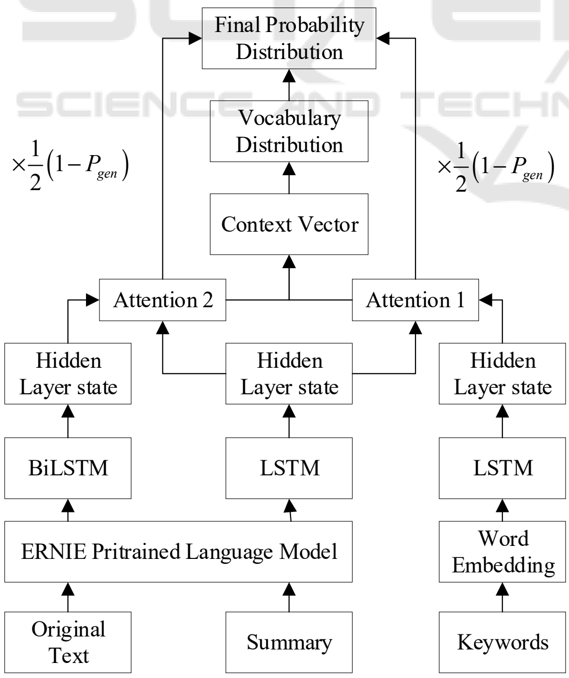

Figure 2: GMDPK network model.

The GMDPK model incorporates an ERNIE (Sun et al., 2019) pretraining language model in the word embedding layer, enhancing the representation of Chinese text syntax structure and entity phrases. The introduction of this model enables more accurate and targeted representation of Chinese semantics.

Our baseline model is the pointer generation network model proposed by (See et al., 2017) The single pointer network only selects copied words from the vocabulary or original text, while ignoring the strong semantic nature of text keyword information. We add a keyword pointer on the basis of the single pointer to achieve the effect of copying words from keywords.

We input the word vectors of the generated keywords $\mathrm{Y_i}$ from the previous section and the hidden layer state from the previous time step into the LSTM network in sequence to obtain the hidden layer state at the current time step.

$$S_{kw\_t} = LSTM(\mathrm{Y}_{t-1}, \mathrm{S}_{kw\_t-1}) \quad (5)$$

Where S_{kw_t-1} is the output information of the keywords from the previous moment. S_{kw_t} is the output information of the keywords at the current moment. Calculate attention score based on the output information of keywords at the current moment and the hidden layer state of the encoder, it can be understood as the degree to which keywords pay attention to the original text information. The formula is as follows:

$$m_{kw\_i}^{t} = Soft\max(O) \quad (6)$$

$$O = V^T \tanh\left(W_h h_i + W_s S_{kw\_t} + b_{attn}\right) \quad (7)$$

V, W_h, W_s, b_{attn} are learnable parameters.

Next, we use the attention mechanism Attention1 in Figure 2 to calculate the probability of copying words on keywords. The formula is as follows:

$$P_{c\_\mathrm{k}}(\mathrm{w}) = \sum_{i:\mathrm{k}_i=\mathrm{w}} m_{kw\_i}^{\mathrm{t}} \quad (8)$$

Our dual pointer network calculates pointers using the following equation, which is actually a selection gating probability.

$$P_{gen} = \sigma\left(w_{h^*}^T h_i^* + w_s^T s_t + w_X^T X_{\mathrm{bi}} + w_Y^T Y_i + b\right) \quad (9)$$

$w_{h^*}^T, w_s^T, w_X^T, w_Y^T, b$ are learnable parameters.

Finally, the GMDPK network uses $\mathrm{P_{gen}}$ to choose whether to copy words from the original text or keywords, or from the vocabulary, in order to predict the final distribution of words.

$$P(w) = P_{\mathrm{gen}} P_{\mathrm{vocab}}(w) + \frac{1}{2}\left(1 - p_{\mathrm{gen}}\right) P_{co}(w) \quad (10)$$

$$P_{co}(w) = p_{c\_k}(w) + p_{c\_o}(w) \quad (11)$$

Where $P_{c\_o}(w)$ is the probability of copying words from the original text, $P_{vocab}(w)$ is the probability distribution of the vocabulary.

4 EXPERIMENTAL ANALYSIS

4.1 Datasets and Evaluation Indicators

We use the LCSTS dataset, which is a Chinese dataset used for text summarization tasks. This dataset contains 380000 Chinese news texts and their corresponding abstracts for training and evaluating the performance of text summarization models.

We use ROUGE-N and ROUGE-L evaluation metrics to evaluate the performance of text summarization models. ROUGE-N evaluates the quality of abstracts by comparing the co-occurrence information of n-grams in automatic and manual summaries. ROUGE-L measures the matching degree of the longest common subsequence between automatic and manual abstracts.

4.2 Experimental Parameter Settings

The model in this article was constructed using the Pytorch deep learning framework and trained on NVIDIA 2080ti GPU with the system version of Ubuntu 16.04. The classic pointer generation network was used as the baseline model.

4.3 Experimental Results and Analysis

The LCSTS dataset is tested and compared with other baseline models. The experimental results are shown in the table 1.

In the baseline model, RNN (Mikolov et al., 2015) is the most basic seq2seq architecture model with RNN as the encoder and decoder. CopyNet (Gu et al., 2016) is an attention based seq2seq model with copy mechanism. PGN is the baseline model we use: pointer generation network. BERT-PGN (Yin et al., 2020) is a pointer generation network based on BERT pretraining model. KAPO (Huang et al., 2020) is to add keyword information on a single pointer generation network. GMDPK is our model based on dual pointer network fusion of keywords.

Table 1: Rouge score comparison of different models.

Model	ROUGE-1	ROUGE-2	ROUGE-L
RNN	21.9	8.9	18.6
CopyNet	34.4	21.6	31.3
PGN	36.24	19.16	32.90
BERT-PGN	37.78	20.61	34.30
KAPO	38.91	21.56	35.54
GMDPK	40.6	22.7	36.3

On the Chinese dataset LCSTS, the ROUGE-1, ROUGE-2, and ROUGE-L scores of this GMDPK were significantly improved compared to the baseline model. Compared with the KAPO model, GMDPK has improved ROUGE-1, ROUGE-2, and ROUGE-L by 1.69% 1.14% and 0.76%, respectively. Our results have been improved, indicating that our GMDPK has learned keyword information and generated abstracts that better fit the original text content.

Table 2 selected an LCSTS test sample. As shown in Table 2. the baseline model PGN network failed to extract key information from the source text, resulting in the generated abstract deviating from the original meaning of the text. The KAPO model extracts two important keywords, such as "深圳" and "司机", and the generated abstract is more complete and factual compared to the PGN model, but far from the reference abstract. In contrast, most of the keywords generated by the MTT in GMDPK appear in the reference abstract, such as "司机", "死", "伤", "深圳", and "赔偿". The generated summary does not contain any extra information and accurately identifies the key information in the original text.

Table 2: Examples of summaries and keywords generated by different models.

Source	一辆小轿车，一名女司机，竟造成 9 死 2 4 伤日前，深圳市交警局对事故进行通报：从目前证据看，事故系司机超速行驶且操作不当导致。目前 24 名伤员已有 6 名治愈出院，其余正接受治疗，预计事故赔偿费或超一千万元。
Reference	深圳机场 9 死 24 伤续：司机赔偿超千万
KAPO	通报 司机 证据 伤员 接受 治疗
MTT	司机 死 伤 深圳 出院 赔偿
PGN	交警局通报司机超速行驶伤员治疗
KAPO	深圳交警局通报司机超速行驶操作不当伤员治疗
GMDPK	轿车司机造成 9 死 24 伤事故赔偿千万

5 CONCLUSIONS

This paper proposes a text summarization model based on dual pointer network fused with keywords to address the high semantic nature of keywords in the original text. Firstly, we design a keyword extraction module based on topic perception and title orientation, it enriches semantic features by mining potential themes, and uses highly summarized and valuable information in the title to jointly guide keyword generation. Secondly, we added the ERNIE pre trained language model to enhance the representation of Chinese text syntax structure and entity phrases. Finally, we added a keyword information pointer to the original single pointer generation network, forming a dual pointer network to fully utilize the extracted key word information. The research results on the LCSTS dataset show that compared with other current methods, our method generates less redundant summary information, contains more critical information in the source text, has better readability, and has achieved improvement in the ROUGE evaluation index.

REFERENCES

Sutskever I, Vinyals O, Le Q V 2014 Sequence to Sequence Learning with Neural Networks Preprint arXiv:1409.3215.

Rush A M, Chopra S, Weston J 2015 A Neural Attention Model for Abstractive Sentence Summarization Preprint arXiv:D15-1044.

Over P, Dang H, Harman D 2007 DUC in context Information Processing & Management, vol. 43, no. 6, pp. 1506-1520 Preprint DOI:10.1016/j.ipm.2007.01.019.

Napoles C, Gormley M, Durme B V 2012 Annotated Gigaword Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction.

Nallapati R, Zhou B, Santos C N D, et al. 2016 Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond Preprint arXiv:K16-1028.

Hu B, Chen Q, Zhu F 2015 LCSTS: A Large Scale Chinese Short Text Summarization Dataset Computer Science:2667-2671 Preprint DOI:10.1109/GLOCOM.2007.506.

See A, Liu P J, Manning C D 2017 Get To The Point: Summarization with Pointer-Generator Networks Preprint arXiv:P17-1099.

Wan X, Yang J, Xiao J 2007 Towards an Iterative Reinforcement Approach for Simultaneous Document Summarization and Keyword Extraction Meeting of the Association of Computational Linguistics.

Xu W, Li C, Lee M, et al. 2020 Multi-task learning for abstractive text summarization with key information guide network Journal on Advances in Signal Processing, vol. 2020, no. 1 Preprint DOI:10.1186/s13634-020-00674-7.

Li H, Zhu J, Zhang J, et al. 2020 Keywords-Guided Abstractive Sentence Summarization Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 5, pp. 8196-8203 Preprint DOI:10.1609/aaai.v34i05.6333.

Akuma S, Lubem T, Adom I T 2022 Comparing Bag of Words and TF-IDF with different models for hate speech detection from live tweets International Journal of Information Technology, vol. 14, no. 7, pp. 3629-3635 Preprint DOI:10.1007/s41870-022-01096-4.

Sun Y, Wang S, Li Y, et al. 2019 ERNIE: Enhanced Representation through Knowledge Integration Preprint arXiv:1904.09223.

Mikolov T, Martin Karafiát, Burget L, et al. 2015 Recurrent neural network based language model Interspeech, Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September Preprint DOI:10.1109/EIDWT.2013.25.

Gu J, Lu Z, Li H, et al. 2016 Incorporating Copying Mechanism in Sequence-to-Sequence Learning Preprint arXiv:P16-1154.

Yin X, May J, Zhou L, et al. 2020 Question Generation for Supporting Informational Query Intents Preprint arXiv:2010.09692.

Huang J, Pan L, Xu K, et al. 2020 Generating Pertinent and Diversified Comments with Topic-aware Pointer-Generator Networks Preprint arXiv:2005.04396.