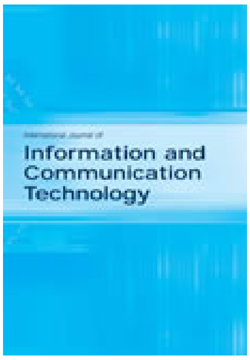
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Automatic summarisation of digital media news based on the transformer model

Xuan Wu, Guoyan Zhang

DOI: 10.1504/IJICT.2025.10071981

Article History:

Received:	01 May 2025
Last revised:	10 May 2025
Accepted:	10 May 2025
Published online:	10 July 2025

Automatic summarisation of digital media news based on the transformer model

Xuan Wu*

Xi'an Jiaotong University City College,
Xi'an 710018, China
Email: xajtdxcsxycb@163.com
*Corresponding author

Guoyan Zhang

Krirk University,
Bangkok 10220, Thailand
Email: BillZhang919@163.com

Abstract: The increasing expansion of digital media content presents a significant difficulty in the field of information processing since it becomes difficult to effectively create accurate and succinct summaries from large news sources. Especially with complex and multimodal news content, traditional news summary generating techniques are sometimes difficult to consider the information coverage; the impact is restricted. This work so suggests, based on the transformer model, an automatic summarising method for digital media news. While decreasing the development of repetitive content via redundancy penalty factors, sentence vector augmentation and keyword guiding techniques help to more precisely capture the relevant information in the news. The strategy suggested in this work greatly beats the conventional summary generating model in the ROUGE series of metrics. New technical solutions and value references for automated processing and intelligent generation of digital media news are presented by this work.

Keywords: digital media news; automatic summarisation; transformer model; multimodal information; redundancy penalty; optimisation strategy.

Reference to this paper should be made as follows: Wu, X. and Zhang, G. (2025) 'Automatic summarisation of digital media news based on the transformer model', *Int. J. Information and Communication Technology*, Vol. 26, No. 25, pp.1–18.

Biographical notes: Xuan Wu received her Master's degree from Shaanxi Normal University in 2020. She is currently a Lecturer at Xi'an Jiaotong University City College. Her research interests include oral communication and new media communication.

Guoyan Zhang received his Doctor's degree from Krirk University, Thailand, in 2024. He is currently a Lecturer at the International College of Arts, Krirk University. His research interests include the theory and practice of broadcasting and hosting, business and corporate communication, media production, and professional media practice.

1 Introduction

Rapid information technology development has made digital media news a major source of information for individuals (Szymkowiak et al., 2021). But the flood of data presents readers of news with a difficult task. An urgent issue now is how fast and precisely one may access important data from the massive news repository. A key tool for addressing this issue, automatic summarisation generating technology has attracted a lot of interest lately. Automatic summarisation technology provides succinct summaries to help readers quickly acquire basic knowledge and enhance information consumption efficiency through the analysis and improvement of news material. Particularly in the realm of digital media news, automatic summarisation technology not only facilitates the fast information sharing but also supports the optimisation of intelligent systems including search engines and news recommendation systems.

As deep learning technology advances, neural network-based automatic summarisation techniques have increasingly become a focus in research. Among them, transformer model has evolved into a major instrument in the field of natural language processing thanks to its strong language modelling capacity and self-attention mechanism (Han et al., 2022). Apart from efficiently capturing the long-distance relationships between sentences, transformer model can also adaptably manage other text generating chores. This work intends to present a transformer model-based novel automatic summarisation approach for digital media news. By means of thorough investigation of news data and the benefits of the transformer model, this work not only maximises the effect of news automatic summarisation but also enhances the quality and variety of the produced summaries.

Compared to state-of-the-art transformer-based summarisation methods, our approach introduces several innovations. Firstly, we incorporate multimodal information to enrich the context understanding of the model. Secondly, we implement a redundancy penalty mechanism to reduce repetitive content in summaries. Additionally, we enhance the model's ability to capture relevant information through sentence vector augmentation and keyword guiding techniques. These improvements collectively enhance the quality and diversity of the generated summaries.

2 Related work

Automatic summarisation methods seek to produce succinct summaries from copious of material in order to enable users to quickly access fundamental knowledge (Singh et al., 2024). The several approaches of generation help one to classify automatic summarisation into two categories: generative and extractive summarising. Usually employing certain heuristic algorithms or conventional machine learning techniques, including TF-IDF, support vector machines (SVMs), and so on, extractive summarising methods essentially produce summaries by choosing significant sentences or phrases from the original text (Hadizadeh et al., 2024). These techniques have clear operation and great efficiency; but their produced summaries sometimes lack verbal fluency and contextual coherence. Consequently, academics have focused more recently on generative summarising techniques, which can provide more natural and varied summaries free from depending on the sentence structure of the source text.

Automatic summarisation approaches have progressively taken front stage as deep learning techniques mature. Mostly depending on recurrent neural networks (RNN) and long short-term memory networks (LSTM), the first generative models. Although these models may reasonably capture the contextual information of sequential data, when processing long texts they still suffer from computational inefficiencies and gradient disappearing. Particularly the transformer model, scholars have suggested solutions based on the attention process to solve these challenges. Transformer is a significant advance in automatic summarisation generation technology, which can effectively capture long-distance dependencies in the text by means of the self-attention mechanism, so enhancing the quality of the summaries and the speed of generating. Transformer not only increases in accuracy but also its parallel computing capacity makes large-scale data processing conceivable unlike conventional RNN and LSTM models (Abed, 2024).

Regarding particular uses, various additional deep learning techniques have been extensively applied for news summarisation creation in addition to transformer. For instance, automatic summarisation techniques grounded in graph neural networks (GNN) are progressively attracting interest. By means of the connection between nodes and the information transfer mechanism, GNN is enable to depict text as a graph structure, so improving the contextual understanding of text (Yang et al., 2021). GNN has a great expressive potential and better model inter-sentence interactions in text than conventional sequence-based models.

Recently investigated are the use of generative adversarial networks (GAN) in summarisation creation. GANs can maximise the generating quality of generative models, lower the overfitting issue, and improve the diversity and naturalness of produced summaries by including an adversarial training mechanism for generators and discriminators. Although the use of GAN in automatic summarisation is yet in the exploratory stage, its potential surely offers a fresh path for study in this domain (Fallahian et al., 2024).

With the ongoing advancement of deep learning technology, several new algorithms progressively remove the constraints of conventional approaches and support the growth of autonomous summarising generating technology. Particularly in the field of news summarisation creation, transformer and its variants' performance has exceeded conventional approaches and grown to be the standard of present study. Still, the profession has a lot of work ahead to raise the accuracy, readability, and variety of the summaries.

This work reflects mostly in the proposal of a transformer-based automatic summarisation approach for digital media news, which integrates multimodal information, redundancy penalty factors, and optimisation strategies. This work introduces sentence vector improvement and keyword guidance mechanism to improve the ability of the model to grasp important information unlike conventional approaches. Furthermore, the adoption of the optimisation technique guarantees the conciseness and information density of the produced summarisation and thus helps to minimise the superfluous information in them. Particularly in terms of semantic consistency, information breadth and fluency, these developments help the model to greatly raise the quality of automatic summarisation.

3 Automated summarisation method based on transformer

3.1 Overview of the transformer model

Based on the self-attention mechanism, the transformer model computes the interactions among items in the input sequence to capture long-range interdependence. Self-attention's basic computational mechanism is based on query, key and value's interaction (Lv et al., 2022). First evaluated using the dot product of query and key, the correlation between the elements is then normalised by softmax to get the attention weights; subsequently, the weighted output is produced by multiplying it with the value matrix. Its mathematical portrayal is:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

where d_k is the key dimension; Q, K, and V are the query, key, and value matrices respectively. This approach allows the transformer to create a new representation for every position depending on the weighted outcome of all other positions in the input sequence.

Figures 1 and 2 offer an illustration of the self-attention mechanism. Figure 1 displays the key to query attentional weights. Figure 2 illustrates the output following weighting depending on the attentional weights; these weights reflect the correlation between various points in the input sequence. Each position in the input sequence is weighted to produce this output, which serves to provide a fresh representation of every location (Avsec et al., 2021).

Figure 1 Weight distribution of the self-attention mechanism (see online version for colours)

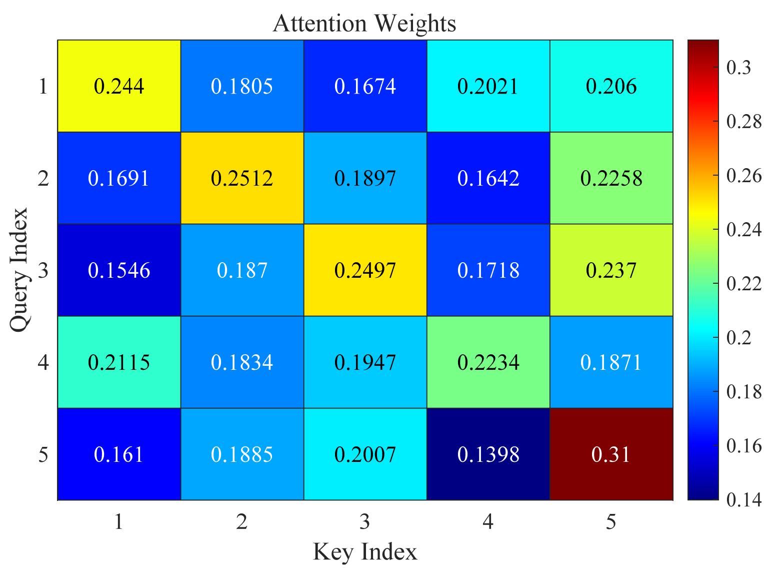
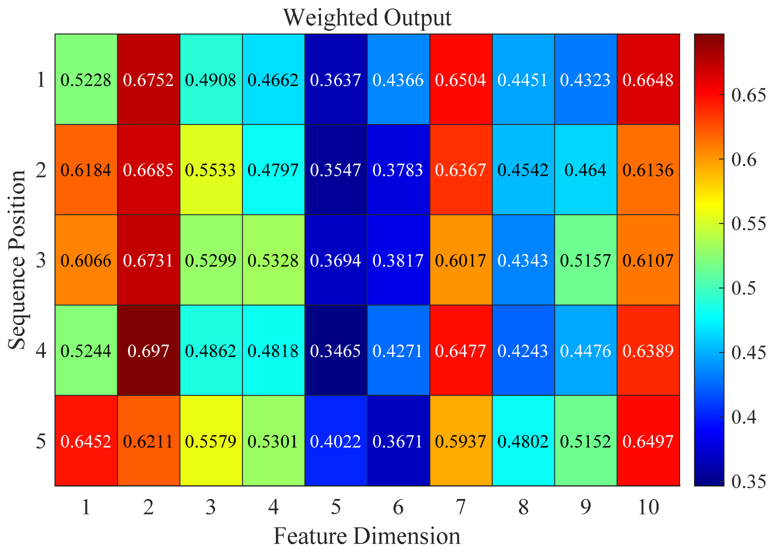


Figure 2 Weighted output of the self-attention mechanism (see online version for colours)



Important visual aids for grasping how the transformer model operates are these two pictures, which alternately depict the two main stages of the self-attention mechanism: computing the attention weights and producing the weighted output.

Transformer is built generally using several encoders and decoders. Two major components comprise every encoder layer: a feed-forward neural network and a self-attention layer (Gao et al., 2022). Via a self-attention mechanism, the inputs capture internal dependencies; these are subsequently handled by a feed-forward network. Every layer's inputs are aggregated with its outputs via residual connections, which layer-wise normalise to produce the final output. Calculated for the input X of a certain layer is the output Y as follows:

$$Y = \text{LayerNorm}\left(X + \text{SubLayer}(X)\right) \tag{2}$$

where SubLayer(X) is a sublayer action of the current layer, like feedforward networks or self-attention. This interaction among layers lets every layer improve the input information representation.

Calculating the encoder-decoder attention layer follows this formula:

$$\text{Attention}_{\text{enc-dec}}(Q, K, V) = \text{softmax}\left(\frac{QW_Q K^T}{\sqrt{d_k}}\right)V \tag{3}$$

where QW_Q is the decoder query matrix multiplied with a learnable matrix W_Q, the output is used for dot product with the encoder's keys. This dot product result is softmax normalised and subsequently multiplied with the value matrix V to provide the final weighted output, same as the self-attention method. In this sense, the decoder can make use of the contextual information given by the encoder to guarantee that the produced textual content is quite consistently highly associated with the input sequence.

Transformer presents a multi-head attention technique to improve the expressive capability of the model. Multi-head attention generates its last output by splicing the outputs of every head and subsequently applying a linear transformation (Yan et al., 2022). Its computing formula is:

$$\text{MultiHead}(Q, K, V) = \text{concat}\left(Z_1, Z_2, ..., Z_h\right)W^O \tag{4}$$

where Z_1, Z_2, …, Z_h are the outputs of every header; concat(•) indicates the stitching of these outputs; W^O is the linear transformation matrix of the outputs. By means of this method, the transformer may interpret the incoming data more holistically and acquire more fine-grained features.

To offset this, though, the transformer model lacks natural capacity to interpret sequential data; hence, positional encoding must be included. By creating unique positional information for every element in the input sequence, positional encoding guarantees that the model comprehends the relative location of items (Liang and Huang, 2024). Usually, position encoding is accomplished by combining sine and cosine functions, that is, computed as:

$$PE_{(pos, 2i)} = \sin\left(\frac{pos}{1,000^{2i/d}}\right) \tag{5}$$

$$PE_{(pos, 2i+1)} = \cos\left(\frac{pos}{1,000^{2i/d}}\right) \tag{6}$$

where pos is the sequence's position; i is the dimension index; d is the dimension of the position encoding. Transformer can properly consider the relative order of elements when processing sequences by means of position encoding.

Transformer's feedforward neural network finally comprises of two linear transformations and a ReLU activation function for non-linear transformation of the self-attentive output. The feedforward network generates as its output:

$$F(X) = \max\left(0, XW_1 + b_1\right)W_2 + b_2 \tag{7}$$

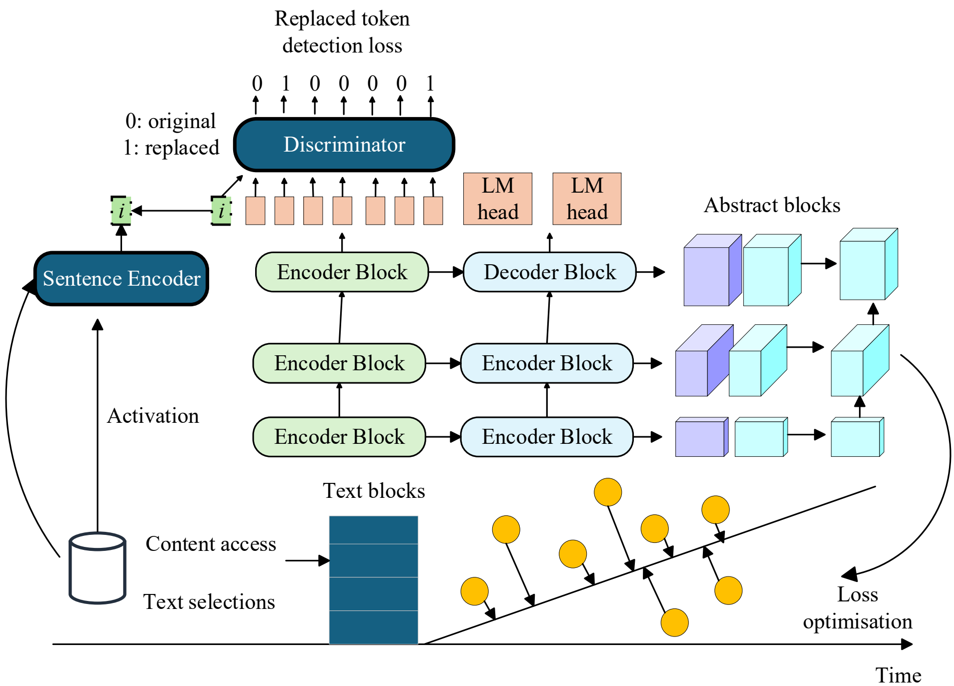
where max(0, •) signifies the ReLU activation function, W_1 and W_2 are the weight matrices of the linear transformation, b_1 and b_2 are the bias terms. Transformer can improve the knowledge and presentation of the input data even more by means of the processing of this feedforward network (Touvron et al., 2022).

By means of these methods, the transformer model not only efficiently executes parallel computing but also successfully records long-range dependencies in the input sequence. Transformer has become one of the most often used models nowadays since these characteristics allow it to thrive in a broad spectrum of natural language processing chores.

3.2 Implementation of automatic summarisation methods for digital media news

Three primary modules are proposed in this study to get high-quality automatic summarising for digital media news scenarios: news text preprocessing, upgraded transformer-based summarisation generating, and summarisation post-processing and optimisation module. See Figure 3 to introduce phrase vector augmentation, keyword guiding mechanism, redundancy penalty factor and other approaches based on the original transformer model to make the model more relevant and practical in the news environment.

Figure 3 Digital media news automatic summarisation method (see online version for colours)



3.2.1 *News text pre-processing module*

News text shows the traits of clear organisation, rich content and varying durations under the background of digital media (Al-Rawi, 2019). Before input modelling, the preprocessing module performs the main duties of formatting, semantic abstraction, and structural expression to ensure the processing effect of transformer model on this kind of text. Three processing techniques comprise the news text preprocessing module developed in this work: sentence vector generation, local structural position encoding, and inter-sentence semantic truncation strategy.

First, this work presents the sentence vector improvement mechanism to solve the issue of different structural but differing semantic levels between sentences in news text by averaging the word vectors in a sentence to get the general semantic representation:

$$s_i = \frac{1}{n_i} \sum_{j=1}^{n_i} w_{ij} \tag{8}$$

where s_i marks the vector of the i^{th} sentence; w_{ij} is the pre-trained word vector of the j^{th} word in the sentence; n_i is the sentence's word count. This approach captures sentence-level semantic characteristics and minimises the complexity of the model input by evenly mapping varying length sentences to the same vector space.

Second, this work local structural improvements to the positional encoding method considering transformer's lack of sequence modelling capacity. To generate a locally position-enhanced position encoding, a paragraph-aware word is included to the standard position encoding mechanism:

$$PE_{(pos,\,2i)} = \sin\left(\frac{pos}{1{,}000^{2i/d_{\text{model}}}}\right) + \lambda \cdot \tan h(\delta \cdot pos) \tag{9}$$

$$PE_{(pos,\,2i+1)} = \cos\left(\frac{pos}{1{,}000^{2i/d_{\text{model}}}}\right) + \lambda \cdot \tan h(\delta \cdot pos) \tag{10}$$

where λ and δ are hyperparameters one can learn or set to control the relative positioning information of words inside a paragraph. This enhanced system not only preserves transformer's original global perspective capacity but also brings modelling preferences for local structures (e.g., introductions and conclusions).

This work proposes an inter-sentence semantic similarity-driven trimming approach to finally prevent incorrect information interference from extended news. Cosine similarity helps one to determine the meaning overlap between adjacent sentences:

$$\text{sim}\left(s_i, s_j\right) = \frac{s_i \cdot s_j}{\left\|s_i\right\|\left\|s_j\right\|} \tag{11}$$

Consecutive sentences can be combined as one logical paragraph or trimmed before summarisation production to increase the information density and improve the efficiency of later generation when their similarity exceeds the given threshold θ.

In essence, the preprocessing module offers the summarisation generating model with triple assurance of semantic integration, positional structure awareness and redundancy control, which greatly enhances the quality of fundamental modelling in the task of digital media news summarising.

3.2.2 Automatic summarisation generation module

Based on the encoder-decoder structure of the transformer model, the central module of this work generates news summaries automatically. By use of many optimisation techniques, including a keyword guidance mechanism, a redundancy penalty mechanism, and a combined training loss function, the module enhances the quality and variation of the summaries. Assuming X as the input word embedding matrix, the encoder's output representation H can be computed with this equation:

$$H = \text{Encoder}(X) \tag{12}$$

Then, using the contextual representation produced by the encoder, the decoder gradually forecasts the words of the summarisation (Du et al., 2020). The decoder not only depends on the produced words at every level but also bases the next prediction on the global semantic information output by the encoder. The decoder's formula is:

$$P\left(y_t \,\middle|\, y < t, X\right) = \text{softmax}\left(W_o \cdot \text{Decoder}(y < t, H)\right) \tag{13}$$

The decoder creates the individual words of the summarisation constantly through this procedure until the whole summarisation is produced.

The model adds a keyword guidance mechanism to raise the quality of the summarisation by means of higher attention weight of the keywords (Li et al., 2023). Particularly, the model dynamically changes the weights depending on the set of keywords *KW* using the following method to estimate the attention:

$$\tilde{\alpha}_i^{(t)} = \tilde{\alpha}_i^{(t)} + \gamma \cdot \mathbb{I}\left(x_i \in KW\right) \tag{14}$$

where γ is a moderating coefficient; $\alpha_i^{(t)}$ is the focus on the i^{th} input word at step t; $\mathbb{I}(\cdot)$ is an indicator function showing whether two successive produced words are the same. This approach helps the model to concentrate on the important facts in the news thereby guaranteeing a more accurate and succinct description.

Redundant information not only compromises the concision of the summarisation but also makes it less readable (Mridha et al., 2021). Thus, by adding a redundancy penalty term, this work limits the model to prevent the emergence of duplicates in the generating process. Definition of the redundancy penalty word is:

$$L_{\text{rep}} = \eta \cdot \sum_{t=2}^{T} \mathbb{I}\left(y_t = y_{t-1}\right) \tag{15}$$

Reducing duplicate generation and enhancing the quality and diversity of summarisation is made possible by punishing unnecessary information in the generating process.

This work also uses a joint training technique to balance the losses of the creation and extraction activities, therefore optimising the quality of the summaries (Zhou et al., 2022). The ultimate total loss function is formed by aggregating the losses of the generating and extraction activities; it is stated as:

$$L_{\text{gen}} = -\sum_{t=1}^{T} \log P\left(y_t \middle| y < t, X\right) \tag{16}$$

For the extraction task, the loss function is defined as:

$$L_{\text{ext}} = -\sum_{i=1}^{n} y_i^{\text{ext}} \log \hat{y}_i^{\text{ext}} \tag{17}$$

Three elements define the final joint loss function: generation, extraction and redundancy penalty; it is stated as:

$$L_{\text{total}} = L_{\text{gen}} + \lambda_1 L_{\text{ext}} + \lambda_2 L_{\text{rep}} \tag{18}$$

where each loss is regulated by weight hyperparameters λ_1 and λ_2.

While preserving the diversity and concision of the produced summaries, the model is able to concentrate more on important information while creating summaries and minimise redundant production by using the above optimising procedures. These techniques taken together guarantees the quality, concision, and informativeness of the summaries as well as increases the efficiency of automatic summarising production.

3.2.3 Summarisation post-processing and optimisation module

The module's major objectives are to raise the produced summaries' fluency, readability, and informative integrity. The original summarisation produced guides some post-processing to guarantee the consistency and logic of the content. First, to minimise any grammatical mistakes or incomprehensible statements that might have developed throughout the generating process, the created summaries are syntactically fixed and linguistically corrected. The model reorders and polishes the produced phrases using a language model so guaranteeing the readability of the summarisation. Assuming we score every produced sentence y using the language model LM, $S(y)$ of the sentence can be stated as:

$$S(y) = \sum_{i=1}^{T} \log P\left(y_i \mid y < i\right) \tag{19}$$

where T is the sentence's overall length; $P(y_i \mid y < i)$ is the likelihood of producing the word y_i given the antecedent $y < i$; y_i is the i^{th} word in the sentence. This helps the language model to maximise the naturalness and fluency of the produced summaries.

Furthermore, the model employs a redundancy elimination technique to help to eliminate needless repetitions in the produced summaries and hence lower redundancy. Based on the cosine similarity computed between every produced summarisation sentence and another sentence, this approach finds duplicate content. Should a given produced sentence's similarity surpass a predefined level, the algorithm either modifies or rejects it. The cosine similarity computed is:

$$\text{cosine}(a, b) = \frac{a \cdot b}{\|a\|\|b\|} \tag{20}$$

where $\| a \|$ and $\| b \|$ are their paradigms and a and b are respectively the vector representations of the two summarisation phrases. By means of this approach, the duplicates in the summarisation can be efficiently eliminated and the concision of the summarisation may be strengthened.

At last, this work presents a keyword-based content modification approach to help to raise the quality of automatic summarisation even more. The coverage of the generated summarisation material is changed to guarantee that the summarisation can cover all the important information by means of analysis of the keyword matching degree between the keywords in the produced summarisation and the keywords in the original news text. Keyword matching degree has the formula shown below:

$$\text{Keyword-Match}(S, T) = \frac{|S \cap T|}{|S \cup T|} \tag{21}$$

where S is the set of keywords in the produced summarisation; T is the set of keywords in the original text; $| S \cap T |$ is the size of the intersection of the two sets; and $| S \cup T |$ is the size of the concatenation of the two sets. This approach guarantees that the automatic summarisation can best preserve the fundamental information of the original text.

4 Experimental design and analysis of results

4.1 Experimental data

The cnn_dailymail dataset is used for the studies to validate the automatic summarising of digital media news based on the transformer model suggested in this work. As the primary means of information distribution nowadays, digital media news consists of many real-time news reports and information summaries with great timeliness and diversity. Widely utilised in automatic summarising research, particularly for automatic summarisation tasks, this dataset comprises news items and their related manually produced summaries from two prominent news websites, CNN and Daily Mail.

With a training set, a validation set, and a test set, the cnn_dailymail dataset has statistical data of every subset roughly as in Table 1.

Table 1 Statistics of the cnn_dailymail dataset

Dataset	*Number of articles*	*Average article length (words)*	*Average summarisation length (words)*
Training set	287,226	758	56
Validation set	13,368	755	56
Test set	11,490	756	56

Every news piece comes with a brief synopsis averaging 56 words, which quite summarises the main points of the work. Model training uses the training set; validation set is utilised for hyperparameter tuning; test set is used for last model performance evaluation. The dataset's huge and representative character makes it very valuable for assessing the performance of applications in automatic summarising systems.

The CNN and DailyMail dataset is widely used in automatic summarisation research due to its large scale and diverse news articles. It is publicly available under the Creative Commons Attribution-ShareAlike 3.0 Unported License (CC BY-SA 3.0), which allows for its use in research and development of summarisation models. This dataset provides a comprehensive benchmark for evaluating the performance of our proposed method.

The features of digital media news define the complexity and challenge of the summarisation generating task, which offers an appropriate validation scenario for the transformer model suggested in this research. The dataset helps to evaluate the performance and efficiency of the suggested approach in processing big news texts.

4.2 Indicators for model evaluation

An important first stage in the automatic summarising job is assessing the model's generating quality. This work uses the ROUGE series of metrics, a collection of commonly used metrics in text summarising tasks, which can efficiently quantify the similarity between the produced summaries and the reference summaries, therefore enabling a full assessment of the performance of the model. The ROUGE measures specifically ROUGE-1, ROUGE-2, and ROUGE-L, three often used tests for word overlap, binary phrase overlap, and the match of the longest common subsequence, respectively.

- ROUGE-1: calculates the word overlap between the reference and summarisation produced (Gulden et al., 2019). This statistic computes the fraction of identical terms in the produced summarisation and the reference summarisation, therefore assessing the coverage of the summarisation. A higher ROUGE-1 score denotes that the produced summarisation includes more important facts.

- ROUGE-2: calculates the binary phrase overlap between the reference and summarisation produced. ROUGE-2 computes the degree to which binary phrases, that is, two adjacent words, overlap in the produced summarisation and the reference summarisation, therefore evaluating the semantic representational accuracy of the summarisation (Ma et al., 2021). This statistic gauges the generated summaries' degree of phrase structural matching.

  ROUGE-2 measures the overlap of bigrams between the generated summary and the reference summary. A 0.05 improvement in ROUGE-2 indicates a significant enhancement in capturing the contextual relationships and semantic coherence of the summary. This suggests that the generated summary is more likely to retain important details and maintain a higher level of accuracy in representing the original text.

- ROUGE-L: calculates the LCS between the reference and generated summarisation's length. By computing the longest common subsequence between the generated summarisation and the reference summarisation, ROUGE-L evaluates their sequence structural similarity (Rougé et al., 2020). This statistic reflects the fluency of the summaries and the information order in addition to word matching.

These evaluation criteria enable researchers to identify variations in the performance of several models in terms of information coverage, semantic consistency and fluency, therefore reflecting the efficiency of the models in the automatic summarising task.

4.3 Experimental results and analyses

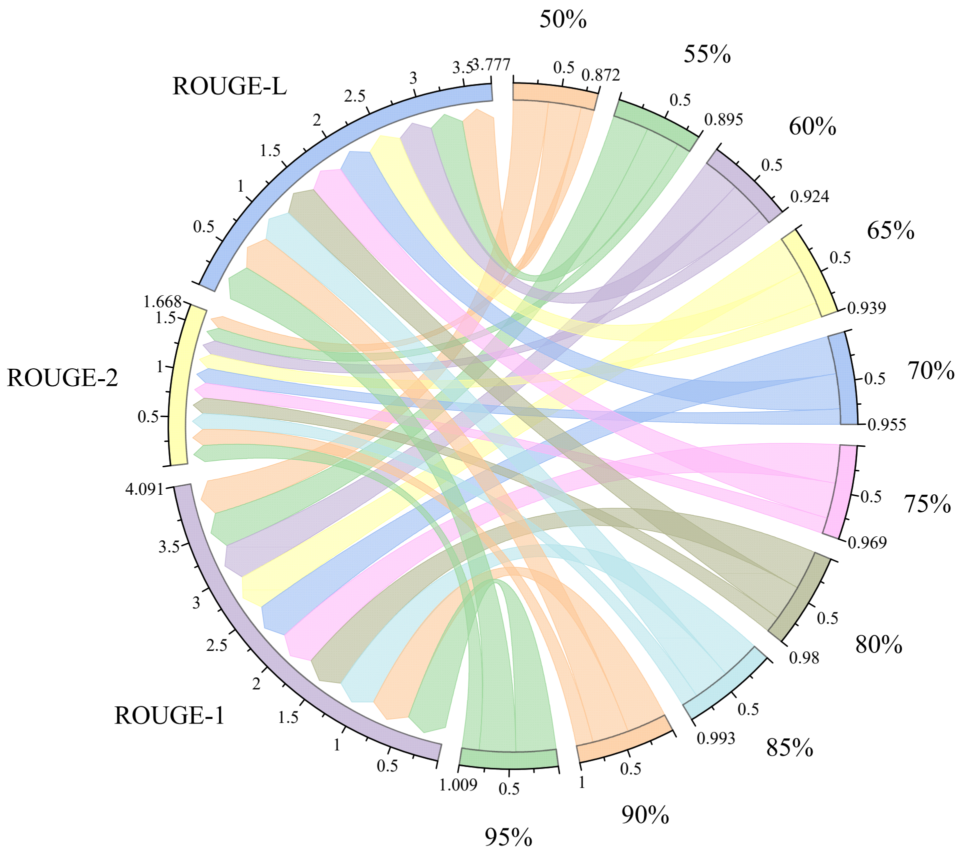
The influence of several training set ratios on the performance of the transformer-based automatic summarising generating model is investigated in first experiment. Specifically, by varying the size of the training set, the model was trained using different ratios of training data (e.g., 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%). The summaries produced at each ratio were assessed. To guarantee the validity and fairness of the studies, the rest of the dataset that including validation and test sets was maintained constant.

Using the cnn_dailymail dataset and selecting various training set ratios, the experiments examine the impact of the training data quantity on the quality of model automatic summarisation. Emphasising coverage, accuracy, and fluency of the summaries, the experiment evaluates the produced summaries using the ROUGE family of metrics.

As the training set ratio rises, the produced summaries from the model progressively get better. Particularly on the ROUGE-1, ROUGE-2, and ROUGE-L measures, the automatic summarisation precision, recall, and fluency the larger the proportion of the training set. This implies that more training data favorably affects the learning and generating capacity of the model.

Figure 4 exhibits the experimental results' presentation:

Figure 4 Effect of proportional division of training set on automatic summarisation (see online version for colours)



Particularly on the ROUGE-2 and ROUGE-L measures, where the rise is more significant, the experimental findings clearly show that the performance of the model progressively improves as the proportion of the training set grows. The improvement in the ROUGE-1 statistic shows that the increase in the training data practically enhances the coverage of the summaries, therefore enabling the model to more fully capture the important information. The rise in the ROUGE-2 measure emphasises the power of the model in binary grammar (2-gram) recall, in which case additional training data improves the model's ability to capture the contextual linkages, hence producing more accurate summaries. Maintaining the fluency and contextual coherence of the summaries depends on the largest common subsequence (LCS), which the rise in the ROUGE-L measure shows the improved capacity of the model to provide. fundamental. Consequently, it can be said that additional training data not only increases the language fluency of the summaries but also boosts their coverage and information correctness.

Conversely, when the proportion of the training set is minimal (e.g., 50% and 55%), the performance of the model is rather poorer on many measures, particularly on the scores of ROUGE-2 and ROUGE-L, which are greatly different from those of the higher proportion of the training set. This implies that inadequate training data can restrict the model's learning of linguistic structures, which influences its knowledge of grammar and contextual linkages during the generating process, and finally results in the produced summaries devoid of accuracy and fluency. Consequently, a reasonable increase in training data, especially in jobs involving a lot of contextual information, such news automatic summarisation, can greatly enhance the generating capacity and summarisation quality of the model. This experimental result supports further model optimisation by verifying the crucial influence of data volume on the performance of deep learning models.

The following experiment will concentrate on assessing the variations in the model's performance on summarisation quality using various assessment criteria when producing summaries in order to better grasp the effect of varied training set ratios. This experiment will enable the analysis of the relevance and benefits of several evaluation criteria in several generating situations.

Emphasising the use of the ROUGE family of metrics (ROUGE-1, ROUGE-2, ROUGE-L), this experiment contrasts the performance of five alternative models while producing summaries. Comparatively analysing the several models helps one to understand the advantages and drawbacks of every model when compiling summaries.

These five models are compared in this experiment:

Model A Based on the basic transformer architecture, a summarisation generating model.

Model B Transformer model with optimisation of positional encoding.

Model C Transformer with keyword steering mechanism combined with sentence vector augmentation.

Model D Transformer model comprising GNN and deep self-attention mechanism.

Model E The model suggested in this work, is a transformer model combining multimodal information, redundant penalty factors, and optimisation techniques for automatic news in digital media summarising.

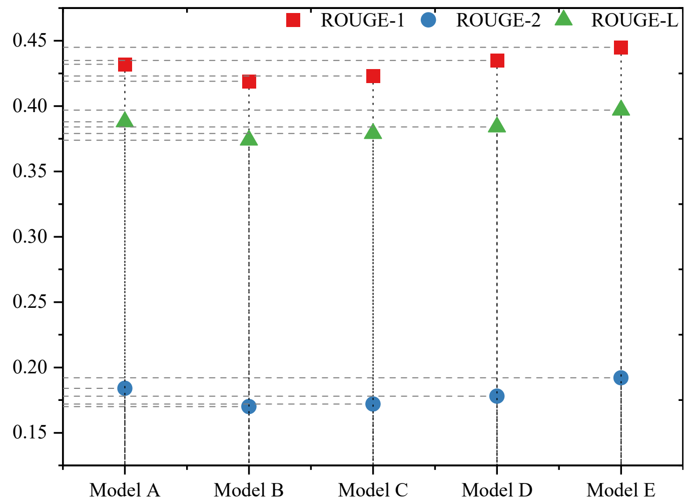
The selection of these models demonstrates the effect of several architectures and approaches on the automatic summarising performance. ROUGE-1 measures word overlap; ROUGE-2 assesses the matching of binary phrases; ROUGE-L measures LCS matching. These evaluation criteria let one evaluate the summaries produced by every model holistically.

Figure 5 illustrates the experimental results' presentation:

Model E has benefits in information coverage and semantic retention as evidenced by the figure, which shows that it performs best in the ROUGE series of measures especially with the highest scores on ROUGE-1 and ROUGE-L. When producing summaries, the transformer model, which incorporates multimodal information, redundancy penalty factors, and optimisation strategies, it showcases more fluency and stronger contextual comprehension.

Higher ROUGE-1 and ROUGE-2 scores for Model A suggest that the typical transformer-based model generates information coverage and accuracy better in general synthesis. Though the spatial coding optimisation increases the structural-awareness of the model, Model B is somewhat lower on ROUGE-1 and ROUGE-2 and fails to considerably raise the quality in generating summaries.

Figure 5 Comparison of different models in automatic summarisation (see online version for colours)



Although Model C scores well on ROUGE-2, implying that it lacks in detail retention and matching of binary phrases when producing summaries, it is optimised for semantic comprehension and keyword guiding mechanisms and performs somewhat weak.

Suggesting that GNN help to improve the model's understanding of the text structure and hence increase the fluency and information coverage of the summaries, Model D, the transformer model that combines the deep self-attention mechanism and GNN, also improves ROUGE-1 and ROUGE-L.

All things considered, the models presented in this work demonstrate a significant benefit by performing the best in the ROUGE series of measures, particularly in producing fluent and semantically coherent summarising. Although both Model D and Model A also help to preserve information integrity, their performance can still be raised by producing more intricate summaries. Conversely, models B and C show quite poor performance in various measures, particularly in ROUGE-2 matching, suggesting still potential for development in model architecture and strategy optimisation.

The company reported a significant increase in quarterly profits, driven by strong sales in emerging markets. The company's quarterly profits surged, fueled by robust sales in emerging markets.

5 Summary and prospects

5.1 Summary of the study

This work suggests a transformer model-based automatic summarising approach using redundant penalty elements and multimodal information to summarise digital media news. Comparative studies confirm the major effects of training set ratio and various model architectures on the quality of automatic summarising generating. The transformer model based on multimodal information and optimisation strategies proposed in this paper performs optimally in the ROUGE series of metrics, especially in terms of information coverage and semantic consistency; the experimental results show that a larger training set ratio can effectively improve the automatic summarisation performance of the model.

This work emphasises the potential of methods including pre-trained language models and GNN by comparing several models, so illustrating the usefulness of various architectures and tactics in automatic summarisation activities. The trial results show that the transformer model including creative approaches can considerably raise the quality of news summaries in digital media, therefore offering fresh technical solutions for useful purposes. Future studies can investigate the combination and expansion of optimisation algorithms to progressively raise the performance and flexibility of the model.

5.2 Research gaps and future work

There are still some flaws even if the transformer-based automatic summarising of digital media news suggested in this paper performs effectively in various studies. First of all, even if the model generates news summaries with greater quality, its generating effect still has to be improved for some news contents with strong domain characteristics and complicated structures. Second, the training set employed in this research is mostly based on public datasets and lacks thorough testing on news data in real-world application environments, thus the adaptability and generalisation capacity in real applications still need to be further verified. Furthermore, even if this work introduces multimodal information and redundant penalty factors, it is still difficult to combine several modal information more effectively when working with multimodal data.

The following will help to better future studies in several spheres. First of all, adding more domain-specific datasets helps the model perform in particular situations. Second, investigating more effective multimodal fusion methods and ways to minimise the negative effect of redundant information on the quality of the summaries helps to optimise the model's structure. At last, combining cutting-edge techniques such deep reinforcement learning to increase the adaptability of the model and intelligent generation in dynamic surroundings (Mohamed, 2023).

It is crucial to address ethical concerns in AI-generated news summaries. Dataset bias can lead to unfair or misleading summaries, so it is essential to use diverse and representative datasets. Additionally, ensuring transparency in the summarisation process and avoiding the spread of misinformation are vital for maintaining the integrity of AI-generated content. Future work should focus on developing methods to detect and mitigate bias in datasets and summaries.

Current multimodal fusion methods face challenges in effectively integrating information from different modalities. One reason is the lack of a unified representation that captures the essence of both textual and visual information. Future research should focus on developing more sophisticated fusion techniques, such as cross-modal attention mechanisms and joint embedding spaces, to better leverage multimodal data. Additionally, exploring the use of advanced models like vision transformers (ViT) in conjunction with text models could provide new insights into multimodal summarisation.

Declarations

All authors declare that they have no conflicts of interest.

References

Abed, A.H. (2024) 'The applications of deep learning algorithms for enhancing big data processing accuracy', *International Journal of Advanced Networking and Applications (IJANA)*, Vol. 16, No. 2, pp.6332–6341.

Al-Rawi, A. (2019) 'Viral news on social media', *Digital Journalism*, Vol. 7, No. 1, pp.63–79.

Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J.R., Grabska-Barwinska, A., Taylor, K.R., Assael, Y., Jumper, J., Kohli, P. and Kelley, D.R. (2021) 'Effective gene expression prediction from sequence by integrating long-range interactions', *Nature Methods*, Vol. 18, No. 10, pp.1196–1203.

Du, S., Li, T., Yang, Y. and Horng, S-J. (2020) 'Multivariate time series forecasting via attention-based encoder-decoder framework', *Neurocomputing*, Vol. 388, pp.269–279.

Fallahian, M., Dorodchi, M. and Kreth, K. (2024) 'GAN-based tabular data generator for constructing synopsis in approximate query processing: challenges and solutions', *Machine Learning and Knowledge Extraction*, Vol. 6, No. 1, pp.171–198.

Gao, C., Zhang, N., Li, Y., Bian, F. and Wan, H. (2022) 'Self-attention-based time-variant neural networks for multi-step time series forecasting', *Neural Computing and Applications*, Vol. 34, No. 11, pp.8737–8754.

Gulden, C., Kirchner, M., Schüttler, C., Hinderer, M., Kampf, M., Prokosch, H-U. and Toddenroth, D. (2019) 'Extractive summarization of clinical trial descriptions', *International Journal of Medical Informatics*, Vol. 129, pp.114–121.

Hadizadeh, Z., Nazari, A. and Mansoorizadeh, M. (2024) 'Enhancing suggestion detection in online user reviews through integrated information retrieval and deep learning approaches', *Journal of Web Engineering*, Vol. 23, No. 3, pp.431–463.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C. and Xu, Y. (2022) 'A survey on vision transformer', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 1, pp.87–110.

Li, X., Li, M., Yan, P., Li, G., Jiang, Y., Luo, H. and Yin, S. (2023) 'Deep learning attention mechanism in medical image analysis: basics and beyonds', *International Journal of Network Dynamics and Intelligence*, pp.93–116.

Liang, H-M. and Huang, C-Y. (2024) 'Integrating non-fourier and ast-structural relative position representations into transformer-based model for source code summarization', *IEEE Access*, Vol. 12, pp.9871–9889.

Lv, Z., Huang, X. and Cao, W. (2022) 'An improved GAN with transformers for pedestrian trajectory prediction models', *International Journal of Intelligent Systems*, Vol. 37, No. 8, pp.4417–4436.

Ma, T., Pan, Q., Rong, H., Qian, Y., Tian, Y. and Al-Nabhan, N. (2021) 'T-bertsum: topic-aware text summarization based on BERT', *IEEE Transactions on Computational Social Systems*, Vol. 9, No. 3, pp.879–890.

Mohamed, M. (2023) 'Agricultural sustainability in the age of deep learning: current trends, challenges, and future trajectories', *Sustainable Machine Intelligence Journal*, Vol. 4, No. 2, pp.1–20.

Mridha, M.F., Lima, A.A., Nur, K., Das, S.C., Hasan, M. and Kabir, M.M. (2021) 'A survey of automatic text summarization: progress, process and challenges', *IEEE Access*, Vol. 9, pp.156043–156070.

Rougé, L., Chiang, N., Steffek, M., Kugel, C., Croll, T.I., Tam, C., Estevez, A., Arthur, C.P., Koth, C.M. and Ciferri, C. (2020) 'Structure of CD20 in complex with the therapeutic monoclonal antibody rituximab', *Science*, Vol. 367, No. 6483, pp.1224–1230.

Singh, L., Sharma, R.K., Saini, N. and Singh, M. (2024) 'Data summerization and voice assistant', *Indian Journal of Image Processing and Recognition (IJIPR)*, Vol. 4, No. 2, pp.9–12.

Szymkowiak, A., Melović, B., Dabić, M., Jeganathan, K. and Kundi, G.S. (2021) 'Information technology and Gen Z: the role of teachers, the internet, and technology in the education of young people', *Technology in Society*, Vol. 65, p.101565.

Touvron, H., Bojanowski, P., Caron, M., Cord, M., El-Nouby, A., Grave, E., Izacard, G., Joulin, A., Synnaeve, G. and Verbeek, J. (2022) 'Resmlp: Feedforward networks for image classification with data-efficient training', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 4, pp.5314–5321.

Yan, W., Zhang, B., Zuo, M., Zhang, Q., Wang, H. and Mao, D. (2022) 'AttentionSplice: an interpretable multi-head self-attention based hybrid deep learning model in splice site prediction', *Chinese Journal of Electronics*, Vol. 31, No. 5, pp.870–887.

Yang, J., Liu, Z., Xiao, S., Li, C., Lian, D., Agrawal, S., Singh, A., Sun, G. and Xie, X. (2021) 'Graphformers: GNN-nested transformers for representation learning on textual graph', *Advances in Neural Information Processing Systems*, Vol. 34, pp.28798–28810.

Zhou, Y., Chen, Z., Wang, B., Li, S., Liu, H., Xu, D. and Ma, C. (2022) 'BOMSC-Net: boundary optimization and multi-scale context awareness based building extraction from high-resolution remote sensing imagery', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, pp.1–17.