Knowledge-Based Systems 227 (2021) 107093

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A hybrid deep generative neural model for financial report generation

Yunpeng Ren [a,b], Wenxin Hu [a], Ziao Wang [a], Xiaofeng Zhang [a,*], Yiyuan Wang [a], Xuan Wang [a]

[a] School of Computer Science, Harbin Institute of Technology Shenzhen, Shenzhen, China
[b] Qianhai Financial Holdings Co., Ltd., Shenzhen, China

ARTICLE INFO

Article history:
Received 20 April 2020
Received in revised form 16 November 2020
Accepted 26 April 2021
Available online 29 April 2021

Keywords:
Financial data mining
Text generation
Natural language generation

ABSTRACT

Generating long macro reports from a piece of breaking news is quite a challenging task. Essentially, this task is a long text generation problem from short text. Apparently, the difficulty of this task lies in the logic inference of human beings. To address this issue, this paper proposes a novel hybrid deep generative neural model which first learns the outline of the input news and then generates macro financial reports from the learnt outline. In the outline generation component, we generate the outline text using the framework of Pointer-Generator network with attention mechanism. In the target report generation component, we generate the macro financial reports by the revised VAE model. To train our end-to-end model, we have collected the experimental dataset containing over one hundred thousand pairs of news-report data. Extensive experiments are then evaluated on this dataset. The proposed model achieves the SOTA performance against both the baseline models and the state-of-the-art models with respect to evaluation criteria BLEU, ROUGE and human scores. Although the readability of the generated reports by our approach is better than that of the rest models, it remains an open problem which needs further efforts in the future.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Text-to-text generation [1] as one of the most significant tasks in natural language generation (NLG) has long been investigated in various domains, e.g., report generation [2], dialogue generation [3–5] machine translation [6,7], and text summarization [8]. Among these sub problems, long text generation from short text is utmost challenging especially for the generation of macro financial reports.

Generally, delicate human efforts are needed to generate macro financial reports especially given a piece of breaking news, which might be quite demanding. However, the generation process of macro financial reports seriously relies on the macroeconomic analysts' ability and thus the quality of generated reports are diversified. To alleviate these aforementioned issues, this work is thus motivated to automatically generate macro financial reports via the proposed deep neural network models.

In essence, this task is a long text generation problem. There exist a good number of successful approaches for text generation such as recurrent neural network (RNN) related based text generation approaches [9,10], variational autoencoder (VAE) based approaches [11] and generative adversarial network (GAN) based approaches [12,13]. However, there exist two challenging difficulties which invalidate most existing approaches. First, the length of the input news is rather short, e.g., " Federal Reserve will establish a facility to facilitate lending to small businesses via the Small Business Administration's Paycheck Protection Program (PPP) ". Whilst the length of the generated reports is usually much greater than that of the input news. This imposes the first difficulty to most existing text generation approaches. Second, the generation of macro financial reports usually involves human beings' intellectual efforts, e.g., inferring and reasoning abilities. Apparently, these challenges remain outstanding problems and thus need more research efforts.

To address aforementioned issues, we intuitively assume that the macroeconomic analysts may first draft an outline and then write reports based on the outline. The outline is believed to well reflect human beings' reasoning content which cannot be directly generated from the short news. Thus, the appropriate working flow should be that we generate the outline for each piece of macro news, and then generate the corresponding macro financial report from the outline. By simulating the working logic of human beings, we propose this hybrid deep generative neural model. The proposed model first learns the outline of the input macro news and then generates macro financial reports from the learnt outline. Without loss of generality, these textual data, e.g., news and reports, are embedded via the skip-gram model. Inspired by [14], the embedded news data are fed into the Bi-LSTM component to train the contextual representation vector

* Corresponding author.
E-mail address: zhangxiaofeng@hit.edu.cn (X. Zhang).

https://doi.org/10.1016/j.knosys.2021.107093
0950-7051/© 2021 Elsevier B.V. All rights reserved.

which is used to learn the latent word probability distribution for generating the outline. To generate the long text report, a decoder component is employed and a global loss function is proposed to minimize the loss from the input news to the generated macro financial reports. To the best of our knowledge, this work is the first attempt to generate domain-oriented long text report. The major contributions of this work are summarized as follows.

- We propose the hybrid deep generative neural model to generate macro financial reports from a piece of short breaking news. The length of the generated report is greater than 200 which is about 5 to 6 times larger than the length of the input text.
- We have collected a large-scale Chinese macro financial reports dataset with over one hundred thousand pairs of news-report data crawled from several popular financial websites. This dataset could be used as the benchmark dataset to evaluate approaches for text summarization, topic model, and text-to-text generation tasks.
- We have performed extensive experiments on this dataset. Both the basedline model and the state-of-the-art approaches, i.e., the Seq2Seq based model and the VAE model are evaluated for performance comparison. The proposed model achieves the SOTA performance against these compared approaches with respect to the evaluation criteria BLEU [15], ROUGE [16] and human scores.

The rest of this paper is organized as follows. Section 2 summarizes the related work of text-to-text generation approaches. Section 3 details the proposed hybrid deep generative neural model. Extensive experiments are evaluated and the corresponding results are illustrated in Section 4 and we conclude the paper in Section 5.

2. Related work

Existing text-to-text generation approaches can be classified into three categories, i.e., RNN related approaches [10,17,18], variational autoencoder (VAE) based approaches [11,19,20], and generative adversarial networks (GAN) based approaches [12,13]. Then, we briefly review these approaches as follows.

In the literature, RNN type approaches have already achieved the state-of-the-art performance over the conventional techniques, e.g., n-gram based model. The RNN model can well capture the contextual information contained in a sentence [21]. Long short-term memory (LSTM) models [22] have been proposed to solve the vanishing gradient problem. Based on RNN models, sequence-to-sequence based approaches have achieved superior model performance in various text to text generation tasks. Cho et al. [23] proposes a neural network architecture with RNNs as a sequence of encoding and decoding component. The encoder embeds the input sequence of variable-length into a fixed-length feature vector, then the decoder maps the vector back to the target sequence of variable-length. To generate long text, a Seq2Seq+Attn model [24] is proposed which allows to search relevant words from the source sentences. Feng et al. [22] proposes a multi-topic aware long short-term memory (MTA-LSTM) network to generate a paragraph-level Chinese essay. The CopyNet model [25] incorporates the copying mechanism into the learning process of Seq2Seq model which achieves a better model performance. Similar attention based approaches can be seen in [26]. Recently, various transformer based approaches [10,27] and pre-trained models [28–30] have achieved the SOTA performance in related tasks.

The variational autoencoder (VAE) models are also widely seen in text generation task. For instance, the RNN-based VAE model [31] is proposed to learn the feature representations of latent variables at the sentence level. The proposed model can explicitly represent the holistic properties of sentences such as style, topic, and high-level syntactic features. An inference network [32] is proposed to apply on the discrete input to estimate the variational distribution. The authors [33] further models the input text as a discrete latent variable under the variational autoencoding framework. On top of the standard VAE model, a hybrid architecture [34] is proposed that blends fully feed-forward convolutional and deconvolutional components to generate a long sequence of text. Yang et al. [35] proposes a topic-to-easy generation approach by using the prior knowledge. Also, the conditional VAE model [36] is considered as the state-of-the-art approach in this task. The CVAE employs a shared attention layer for both encoder and decoder to learn better feature representations of coherent sentences. Later, a multi-pass hierarchical CVAE [37] is proposed for automatic storytelling. Note that the CVAE-type models are generally resolved through the ELBO optimization [19].

The original generative adversarial network (GAN) [38] is already widely adapted to various research problems. As the original GAN cannot model discrete variables, [39] proposes to employ Gumbel-softmax distribution for this issue. Then, by modeling the data generator as a reinforced stochastic policy, Yu et al. [40] proposes the sequence generation (SeqGAN) framework to directly perform gradient policy updating rules. Furthermore, RankGAN [41] is proposed to generate high-quality textual descriptions. MaskGAN [42] approach is proposed to optimize the process of generating text and LeakGAN [43] is further proposed to generate long text within 40 words. To reduce labeling cost, [44] proposes the GAN-BERT which extends the BERT-like architecture for modeling unlabeled data in a generative adversarial setting.

Among these aforementioned approaches, the hybrid pointer-generator network [14] and the Seq2Seq+Attn model [24] are the most related works to our approach. To improve the performance of encoder–decoder process, the fixed-length vector is used in the Seq2Seq+Attn model to search more relevant source sentences. Although the Seq2Seq based model can generate long text given the short input text data, the inferring and reasoning content cannot be acquired by these approaches. Moreover, despite some reinforcement learning based approaches [35], the length of generated long text is usually less than 40 words [43], which is not suitable for the generation of macro financial reports.

Therefore, this hybrid deep generative neural model based on pointing sampling technique is proposed to simulate the inferring and reasoning ability of human beings.

3. The proposed approach

In this section, the proposed hybrid deep generative neuralmodel is briefly illustrated as follows. The proposed approach consists of two sub components, i.e., outline generation component and text generation component, as plotted in Fig. 1. For the outline generation component, we first generate the outline text O form the input macro news X. The macro news data X contains m words and both of the report Y and the outline text O contain l words, where l is far greater than m. Both X and Y are embedded into word vectors. The word vectors of X and Y are respectively fed into the bidirectional LSTM (Bi-LSTM) module and the employed LSTM module is used to co-train the outline O. Y is to provide external knowledge or information to generate an appropriate outline O in the model training stage. The attention mechanism is introduced to train the contextual representation vector by emphasizing the important words of the input X. As for the text generation component, the learnt O is used to generate the target report Y' via the general VAE model. Note that Y is

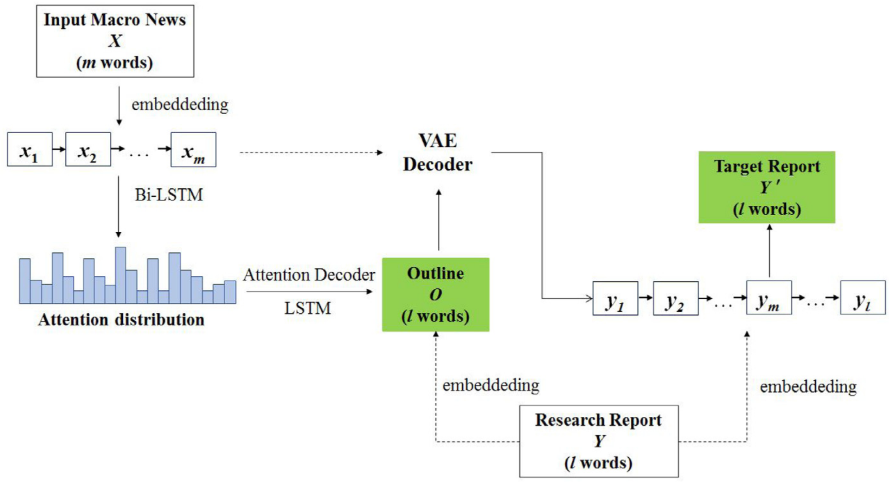

Fig. 1. The framework of the proposed hybrid deep generative neural model.

only provided in the model training process to help to generate a more accurate outline O which is then used to generate the target report Y'. In the testing process, we only provide the macro news X and the pre-trained outline text O as the model input.

3.1. Outline generation component

Inspired by the TA-Seq2Seq model [45] and Pointer-Generator Networks [14], we design this outline generation component as plotted in Fig. 2. Let $X = \{x_1, x_2, \ldots, x_m\}$ represent the macro news data and $O = \{o_1, o_2, \ldots, o_l\}$ represent the outline to be generated. In this component, a sequence of X is first embedded through the widely adopted skip-gram model to capture their contextual information. Then, the embedded word vecters of X are fed into the Bi-LSTM module in a one-by-one manner. The output of the Bi-LSTM is a sequence of hidden states h_t^e, calculated as

$$h_t^e = f_{encoder}(x_t, h_{t-1}^e) \parallel f_{encoder}(x_t, h_{t+1}^e), \tag{1}$$

where $f_{encoder}(x_t, h_{t-1}^e)$ and $f_{encoder}(x_t, h_{t+1}^e)$ represent the hidden states of forward network and backward network, respectively. To generate macro financial report Y, the core semantic content of Y are introduced to co-train the generation of O. To make use of the position embedding information [46], similar to transformer model, Y is marked with $\langle start \rangle$ at the beginning and $\langle end \rangle$ at the end. Then Y is fed into a LSTM model to acquire the hidden states of the decoder. At each step t, the decoder receives the word embedding of the previous word as well as the decoder states s_t^d, computed as

$$s_t^d = f_{decoder}(\hat{o}_{t-1}, s_{t-1}^d). \tag{2}$$

The attention score e_i^t is calculated as

$$e_i^t = v tanh(W_h h_i^e + W_s s_t^d + w_c c_i^t + b_{attn}), \tag{3}$$

where v, W_h, W_s, w_c and b_{attn} are learnable parameters. Similarly, c^t is a cover vector which contains the attention information for all previous moments, calculated as

$$c^t = \sum_{t=0}^{t-1} \alpha^t, \tag{4}$$

where α^t is the attention probability distribution generally calculated using the softmax function, written as

$$\alpha^t = \frac{exp(e_i^t)}{\sum exp(e_i^t)}. \tag{5}$$

The calculated attention probability distribution α^t is then used to weight the encoder hidden states h_t^e to generate the context vector h_t^*, calculated as

$$h_t^* = \sum_i \alpha_i^t h_i^e, \tag{6}$$

The context vector h_t^*, which can be seen as a fixed size representation of the input source information at current step, is then concatenated with the decoder hidden state s_t^d. After that, h_t^* is fed into a two fully connected layers to generate the vocabulary distribution P_V, computed as follows

$$P_V = softmax(V'(V[s_t^d, h_t^*] + b) + b'), \tag{7}$$

where V, V', b and b' are learnable parameters of the two linear layers.

The merit of the original pointer generation is to generate unseen words and this is fulfilled by sampling terms from the vocabulary corpus. Therefore, a soft switch p_{gen} is designed to well balance the generation of a word either from the vocabulary corpus by sampling from P_V, or from the input sequence by sampling from the attention distribution α^t.

At time t, p_{gen} is calculated by the context vector h_t^*, the decoder state s_t^d and the decoder input o_t, given as

$$p_{gen} = \sigma(W_{h^*}^T h_t^* + W_{s^t}^T s_t^d + W_x^T o_t + b_{gen}), p_{gen} \in [0, 1], \tag{8}$$

where vectors $W_{h^*}^T$, $W_{s^t}^T$, W_x^T and scalar b_{gen} are learnable parameters and σ is the sigmoid function. Further, we obtain the final probability distribution over the extended vocabulary set, given as

$$P(w) = p_{gen} P_V(w) + (1 - p_{gen}) \sum_{t:w_i=w} \alpha^t. \tag{9}$$

If w is an out-of-vocabulary word, the first term $P_V(w)$ becomes 0. Similarly, if w does not appear in the source document, the second term of this Equation is 0. On one hand, Eq. (9) solves the problem of generating out-of-vocabulary words. On the other hand, it makes full use of the attention distribution of the original input text.

To avoid generating repetitive text, we also define a *covloss* function to penalize repeatedly generated words at the same position, calculated as

$$covloss^t = \sum_i min(c_i^t, \alpha_i^t). \tag{10}$$

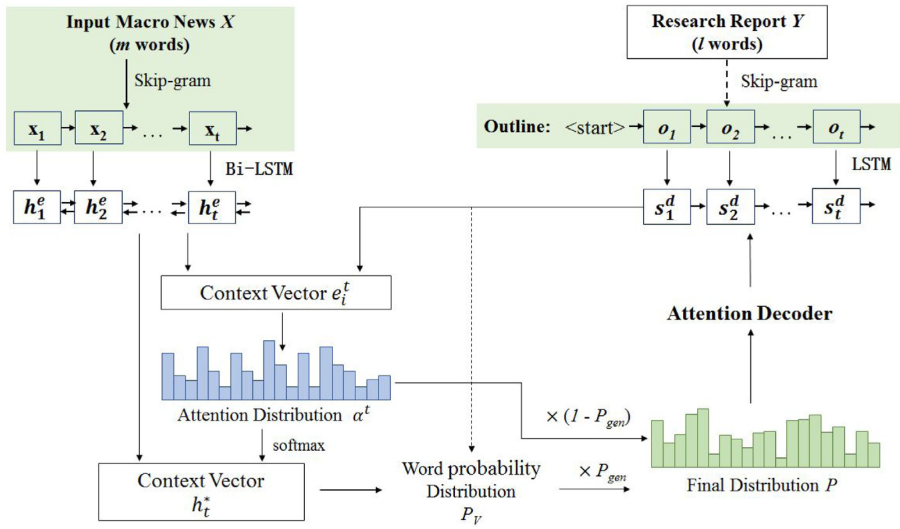

Fig. 2. Details of the proposed outline generator component.

If the generated word at the current moment has been generated, the attention probability of the current word's position will be decreased. Therefore, the corresponding loss function of the outline generation model $L_{outline}$ could be modeled as the superposition of the primary loss function and *covloss* function, calculated as

$$L_{outline} = -logP(w_t^*) + \lambda \sum_i min(c_i^t, \alpha_i^t), \tag{11}$$

where w_t^* is a generated word and λ is a hyper-parameter.

3.2. Text generation component

The text generation component is considered as a decoding process and thus the VAE based module is a natural choice to be the decoding component. To generate text Y', we take both the macro news X and the pre-trained outline text O as the model input, as shown in Fig. 3. Then, we employ the VAE model to learn the posterior distribution of latent variables $P(z|X)$ and samples data from this distribution as depicted in Fig. 1. The $P(z|X)$ can now be rewritten as

$$P(z|X, O) = \frac{P(X, O|z) \cdot P(z)}{P(X, O)}. \tag{12}$$

The log likelihood of sampled data x can be acquired by maximizing the ELBO problem, defined as follows

$$logP(X, O) \geq E_{q(X,O|z)}logP(X, O|z) - KL(q(X, O|z) \parallel P(z)), \tag{13}$$

In this ELBO problem, the first term is the sampled data from $P(X, O|z)$ to calculate the cross entropy loss and the second term, i.e., the KL divergence, is used to ensure that the posterior distribution is close to the prior distribution. In this component, the decoder's hidden state h_t^d is updated as

$$h_t^d = f_{encoder}(\hat{y}_{t-1}, h_{t-1}^d). \tag{14}$$

Therefore, the loss function L_{report} for the text generation component can be written as

$$L_{report} = \sum_{i=0}^{1} -logP(y_i' = y_i|h^d). \tag{15}$$

At last, the overall loss function of the proposed approach can be calculated as the summation of $L_{outline}$ and L_{report}, written as

$$Loss = L_{outline} + L_{report}. \tag{16}$$

4. Experiment

4.1. Dataset and data preprocessing

To evaluate the proposed approach, we first carefully collected a large-scale experimental dataset from several popular Chinese financial Websites, e.g., Sina Finance, Tonghuashun Finance and Eastmoney net. We respectively crawled 69,960, 52,360 and 8,017 pairs of news-report data from each data source. Note that we only collected the macroeconomic news for this experiment.

For data preprocessing, we eliminated the numeric symbols and other special characters from the original news data. To form the news-report pair, the keywords of these macro news are used to best match the corresponding macro financial reports. For each piece of news, we only match one financial report. One sample pair of data is given in Table 1. The left column is the macro news and the right column is the corresponding financial report written by macroeconomic analysts.

To segment the Chinese macro news data, an open source tool ("jieba") is utilized. After word segmentation, the average and median length of the macro news are 28 and 24 words, respectively. Whilst the average and median length of the financial reports are 341 and 331 words, respectively. As the text generation component requires the length of the input data should be the same, we therefore truncated or lengthened the macro news and financial reports to the same length. Now, the length of input data is set to 30 words, and length of financial report is set to 200 words.

To filter out the rare words, we only kept the word in the vocabulary corpus if its term frequency (TF) is higher than 5. After that, the vocabulary corpus, used for generating unseen words, contains 63,782 words and we filter out about 0.4% terms. Moreover, the vocabulary is marked with four tokens, i.e., padding token (PAD), unknown token (UNK), start position token (START) and end position token (END). The PAD is used to fill into the input of the encoder and decoder. The UNK is used to represent words that are not in the vocabulary corpus which are usually rare but meaningful words, such as the entity name (a company or a person). The START and END are added at the beginning and the end of each report, respectively.

4.2. Experimental settings

To evaluate the proposed approach as well as the baseline models, we randomly chose 90% of the dataset as the training set

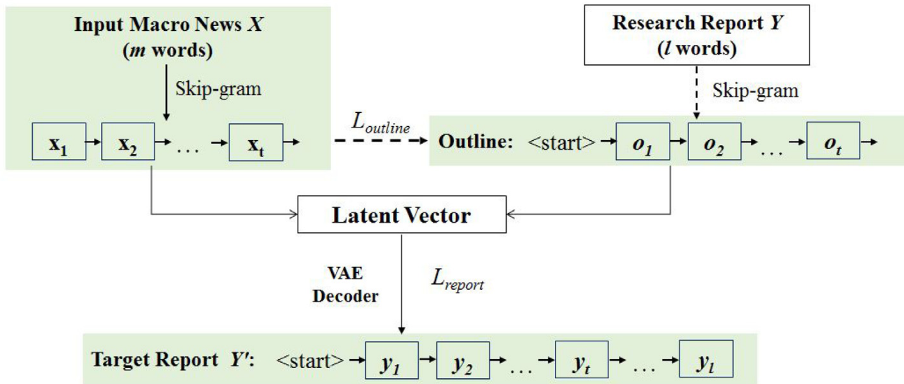

Fig. 3. Details of the proposed text generation component. ($L_{outline}$ and L_{report} represent the loss function of the outline generator component and the text generation component, respectively).

Table 1
A sample of processed macro news and the corresponding financial report.

Macro news	Corresponding macro financial Report
The People's Bank Of China released its financial statistics report for August 2019. The RMB loans increased by RMB 1.21 trillion and the growth rate was 12.4% from a year earlier. Aggregate Financing to the Real Economy increased by RMB 1.98 trillion and the growth rate was 10.7%.	Comments: Reverse cyclical adjustment of may continue to be reflected in financing. Due to the impact of external economic environments, the pressure on the economy has increased. Current economic policy is a two-pronged approach. On one hand, the economy policy helps to improve the medium-term and long-term expectations of macro economy via opening-up and market-oriented reforms. On the other hand, reverse cyclical adjustment will ensure that the economy will not slow down in the short term. In terms of finance, recent measures, such as LPR reform, comprehensive RRR cuts, targeted RRR cuts, and the expanded QFII quotas, have all played a more important role in financial support for real economy. It is expected that under the reverse cyclical adjustment, Aggregate Financing to the Real Economy will be a strong trend in the next few months and the growth rate is expected to reach or even exceed the annual high of 10.9% in June at the end of the year.

and the rest 10% as the testing set. The length of the generated report Y' is required to be 200. For the outline generation component, the one-hot vector is generally adopted for the embedding of text data. Without loss of generality, both X and Y are embedded via the high-dimensional one-hot vector and then mapped into a low-dimensional vector. Before the experiments, we first empirically set several hyperparameters, then we fine-tune these parameters during the model training process to minimize the model loss. The dimension of low-dimensional embedding vector is set to 128. The number of hidden units in the LSTM component is set to 256. The maximum length of encode and decoder are set to 30 and 200, respectively. The learning rate is set to 0.005, the batch size is set to 16 and the weight for repetitive generation loss is set to 1. To improve the efficiency of encoding and decoding process, we use beam search strategies [47] to generate the outline text and the beam parameter is set to 3, which means that the 3 words with the highest probability will be selected out as the candidate words to generate the report. For the text generation component, the VAE based component is to be trained. In the proposed VAE component, the dimension of the hidden variables is set to 120, the dropout rate is set to 0.5 to avoid over-fitting problem, the batch size is set to 16.

4.3. Baseline models

To evaluate the model performance, several baseline and the state-of-the-art approaches, i.e., the Seq2Seq model [48], the Seq2Seq+Attn model [24], the Pointer-Generator model [14] and the VAE model [49] are performed. Details of these approaches are illustrated as follows.

- The Seq2Seq model is considered as a baseline model which already achieves a superior performance in various text-to-text generation related tasks, and thus is chosen for the performance comparison.
- The Seq2Seq+Attn model [24] extends the original Seq2Seq model, and is originally proposed for neural machine translation task. In this model, a fixed-length vector is adopted for both encoder and decoder component, and it allows to automatically soft-search relevant words from the input source sentence to predict the next word to be generated, without having to form these parts as a hard segmentation task.
- The Pointer-Generator model [14] is first proposed for text summarization task. In this model, a hybrid pointer-generator network is designed to sample words from the input source text data via the so-called pointing process, and then integrates the coverage mechanism to keep track of generated words to penalize repetitive generation.
- The VAE model is one of the most widely adopted models for the text generation task. It learns the neural network model by minimizing the error between the reconstructed data and the input data. The learnt decoder component, low-dimensional latent variable space, could then be applied to generate high-dimensional data.

4.4. Evaluation metric

For the evaluation criteria about the text generation task, both the objective and subjective criteria are chosen in the experiments including the Bilingual Evaluation Understudy (BLEU), ROUGE and human score. As most Chinese words consist of less than 5 characters, we chose the BLEU-2, BLEU-3 and BLEU-4 scores as the evaluation metric to evaluate the results generated by all approaches. Similarly, ROUGE-1, ROUGE-2 and ROUGE-L are used to evaluate the model performance.

4.5. Evaluation results

The evaluation results of the proposed hybrid deep generative neural model as well as all compared methods are respectively reported in Tables 2–4.

Table 2
Results of our model and compared methods in terms of BLEU criteria.

Methods	BLEU-2	BLEU-3	BLEU-4
Seq2Seq	0.253	0.291	0.102
Seq2Seq+Attn	0.289	0.332	0.112
Pointer-Generator	0.322	0.366	0.142
VAE	0.311	0.233	0.08
Hybrid Deep Generative Neural Model	**0.432**	**0.389**	**0.164**

Table 3
Results of our model and compared methods in terms of ROUGE criteria.

Methods	ROUGE-1	ROUGE-2	ROUGE-L
Seq2Seq	0.032	0.021	0.043
Seq2Seq+Attn	0.066	0.053	0.095
Pointer-Generator	0.133	0.087	0.114
VAE	0.081	0.011	0.023
Hybrid Deep Generative Neural Model	**0.141**	**0.084**	**0.127**

Table 4
Results of our model and compared methods in terms of human score.

Methods	Fluency	Consistency
Seq2Seq	1.54	0.87
Seq2Seq+Attn	2.86	2.31
Pointer-Generator	3.35	3.58
VAE	2.03	2.19
Hybrid Deep Generative Neural Model	**3.62**	**3.69**

For objective evaluation results, Table 2 reports the evaluation results on the BLEU criteria in which our proposed model achieved the highest score. The BLEU-2, BLEU-3 and BLEU-4 scores of our proposed approach are 0.432, 0.389 and 0.164 which are 34%, 6%, 15% higher than the second highest scores of the Pointer-Generator methods, respectively. Table 3 reports the evaluation results on the ROUGE criteria. The Pointer-Generator model is more effective with respect to the ROUGH criterion among other baseline methods. The proposed model tend to achieve higher ROUGE score than the Pointer-Generator model. It is also noticed that for ROUGE-2, the score of the Pointer-Generator is close to our proposed model. Moreover, the ROUGE-1 and ROUGE-L scores of our model are respectively 6%, 11% higher than that of the Pointer-Generator model. From the results about the objective evaluation criteria, it is obvious that the proposed approach already achieves a better model performance.

For subjective evaluation results, we chose the human scores to further evaluate the fluency and consistency of the generated reports. In this experiment, we randomly chose 1000 macro financial reports and manually annotated these reports. This human score annotation is also a common practice in NLP related task. Table 4 reports the evaluation results on the human scores of all approaches. The higher this value, the better the model performance. From this table, it is well noticed that the hybrid deep generative neuralmodel achieves the best human score values. Note that the second best model is the Pointer-Generator model which is also consistent with our expectation.

To summarize, it is well noticed that, from the evaluation results of the BLEU, ROUGE and human score, the model performance of the proposed hybrid deep generative neuralis superior to the baseline model as well as the state-of-the-art approaches. Both these objective and subjective evaluation results could well verify the effectiveness of the proposed approach in generating long text from short text. For subjective evaluation, we also report one generated outline text O and one target report Y' extracted from the testing results, given in Table 5, for users to make their own subjective judgement. In the target report, we highlight the correct content in blue color. The logical frame of the outline text O is similar to the report text, but the language coherence and readability are not as good as the target report Y'. Table 5 shows that the coverage ratio of the correctly generated content is acceptable. Furthermore, we compared the generated target reports Y' by our hybrid deep generative neural model and the 4 baseline models, as shown in Table 6. And we highlight the correct content in blue color. The target reports Y' generated by Seq2Seq model and Seq2Seq+Attn model still have a number of repeated words and the logic of the generated content is not satisfying. The VAE model and Pointer-Generator is slightly worse than the proposed model. This result is consistent with that of Table 4 in which the fluency and consistency of the generated reports of our model achieve the best model performance, which further verifies the effectiveness of the proposed model.

4.6. Discussion

In this section, we further elaborate the possible reason why our proposed model outperforms the baseline models as shown from Tables 2 to 4. Although the Seq2Seq model is one of the most important SOTA approaches for text generation task, its performance is not satisfying for this specific task, i.e., generating long text from short text. From Tables 2 to 4, it could be seen that the Seq2Seq model only achieves the lowest score especially in the human score measuring the consistency of the generated text and target text. The possible reason is that the lack of sufficient information of the input data. Therefore, it is really challenging to generate a longer text given such insufficient information. Accordingly, the generated text using these models tend to repeat themselves. However, the Seq2Seq model well suits the situation where the length of input data is equal or greater than that of the output data, such as machine translation and dialogue system. The Seq2Seq+Attn model could significantly improve the quality of the generated text as shown in Table 3. This observation verifies the effectiveness of the attention mechanism with the soft-search strategy. The Pointer-Generator model also employs an attention mechanism and performs well in short-to-long text generation task. Apparently, the attention mechanism truly plays an important role in short-to-long text generation. Compared with the RNN-based neural network models, the VAE-type model directly models the end-to-end generation process. The learnt low-dimensional latent variable layer is thus able to decode (i.e., generate) a high-dimensional output data. This is the possible reason why the VAE type model could also achieve higher evaluation scores. Furthermore, the VAE model with the attention mechanism could emphasize more important words to be generated. The appropriate sequence of the generated words could be guaranteed by the Seq2Seq type model. This is the behind reason why we design our proposed model under the unified framework of combining the VAE structure with the RNN type structure.

5. Conclusion and future work

In this paper, we propose this novel hybrid deep generative neural particularly for generating macro financial reports from short breaking news. This task is a challenging long text generation task. In addition to the existing SOTA long text approaches, the generated text length of our approach could even reach to 200 words whereas the length of the input text is rather limited. To the best of our knowledge, this is among the first attempts to generate the longer text from a comparably short input data. The proposed approach consists of two sub components. In the first component, we generate the outline text using the proposed outline generation component analogous to the framework of the Pointer-Generator network with attention mechanism. In the

Table 5
The generated outline O and report Y' given a piece of macro news X.

Input Macro News X	The European sovereign debt crisis is a "sequelae" phenomenon of the relief policy of financial crisis. The global economy may slow down and fall into a stage of "high debt and low growth" due to the sovereign debt crisis because of the five European countries. Market demand will decrease and the process of going out of the global economic crisis will be prolonged.
Outline text O	US economic growth We expect that the economic growth is slowing down,, the first interest rate hike cycle since the financial crisis, China–US trade friction and other factors and the global financial market, the global financial market may reach a new high since the year. Monetary easing policy.. is for the currency market and the monetary policy of the People's Bank of China is tightening liquidity. The Federal Reserve has announced its monthly economic growth expectations economic growth expectations. Asset risk is: China's central bank has begun to recover after the inflation level. In order to do so, the European indexes have all increased. We believe that the Fed is expected to reach the same period of the year's economic growth; the interbank decline is mainly due to the economic debt crisis and high financial leverage. The interbank decline is mainly due to the economic debt crisis The main reason for the increase in leverage is the interbank market Monthly day Treasury bonds.
Generated Reports Y'	We believe that the expectations of global economic growth may show up in month, which is the highest level since the beginning of the year. The global economic recovery may get a new height since the month of the year. Furthermore, the economic growth of the United States may get a new height recently. Our view is that the US economic growth is the major factor in global economic growth. On the day of the month, the US Federal Reserve announced its annual expectations of economic growth, but we believe that the expectations of the US Federal Reserve may be the expectations to raise interest rates. Monthly exports increased by a percentage point from the previous month. Judging from the economic growth data, the economic growth rate of the United States will be respectively in the year. And the pressure on the global economic recovery will decrease, economic indexes of European will rise. We believe that the expected economic growth of US Federal Reserve is to be in the year. Economic growth is expected to be in month. Annual economic growth will effect by the period of interest rate cycle since the beginning of the year. The global economy will continue to slow down.

Table 6
The input macro news X and the corresponding report Y' generated by our proposed model and four baseline models.

Input Macro News X	The European sovereign debt crisis is a "sequelae" phenomenon of the relief policy of financial crisis. The global economy may slow down and fall into a stage of "high debt and low growth" due to the sovereign debt crisis because of the five European countries. Market demand will decrease and the process of going out of the global economic crisis will be prolonged.
Seq2Seq-report Y'	Our economic data is deteriorating; the second is that economic growth will continue to be around for the year; for the year month) month The "About" was released on the publication day of the year, month, and day of the reporting period.) The government's fiscal problems and fiscal aspects: First, the U.S. Treasury bond yield has risen to-month)......
Seq2Seq+Attn-report Y'	Our economic data will deteriorate further; the second is that economic growth will continue to remain around the year since the year;) The "About" was issued on the year, month, and day of the reporting period, and the "About" "On the issue of government debt and debt. ; Second) The government's fiscal problems and fiscal aspects: the U.S. Treasury bond yield rose to month......
Pointer Generator-report Y'	US economic growth We expect that the economic growth is slowing down,, the first interest rate hike cycle since the financial crisis, China–US trade friction and other factors and the global financial market, the global financial market may reach a new high since the year. Monetary easing policy..is for the currency market and the monetary policy of......
VAE-report Y'	What we think of the growth is that China–US trade relations are also strengthening, and China play a certain role in the relationship between the economy; industrial policies, exchange rates, and economic issues have been the largest growth in the past years. There has been a significant rebound for the first time since, but it has also appreciated significantly since the year, month and day......
Hybrid deep generative neural model-report Y'	We believe that the expectations of global economic growth may show up in month, which is the highest level since the beginning of the year. The global economic recovery may get a new height since the month of the year. Furthermore, the economic growth of the United States may get a new height recently. Our view is that the US economic growth is the major factor in global economic growth. On the day of the month, the US Federal Reserve announced its annual expectations of economic growth, but.....

second component, we generate the macro financial reports from both the generated outline and the input macro news by the revised VAE component. To evaluate the model performance, a large-scale experimental dataset is collected which consists of over one hundred thousand pairs of news-report data. Extensive experiments are then evaluated on this dataset. From the promising experimental results, it is well noticed that the proposed approach significantly outperforms both the baseline and the state-of-the-art approaches w.r.t. the evaluation criteria, i.e., BLEU, ROUGE and human score, which verifies the effectiveness of the proposed approach. In the near future, we will further investigate a better way to generate a much longer text, and try to further improve the readability of the generated reports.

CRediT authorship contribution statement

Yunpeng Ren: Writing, Investigation, Validation, Conceptualization. **Wenxin Hu:** Implementation, Investigation, Validation, Data curation. **Ziao Wang:** Software programming. **Xiaofeng Zhang:** Ideas, Development or design of methodology, Creation of models, Writing, Supervision. **Yiyuan Wang:** Data curation, Software. **Xuan Wang:** Writing, Supervision.

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.

Acknowledgment

This work was supported in part by the National Key Research and Development Program of China under Grant no. 2018YFB1003800, 2018YFB1003804, the National Natural Science Foundation of China under Grant No. 61872108, and the Shenzhen Science and Technology Program under Grant No. JCYJ20170811153507788.

References

[1] Y. Duan, J. Pei, C. Xu, C. Li, Pre-train and plug-in: Flexible conditional text generation with variational auto-encoders., 2019, arXiv: Computation and Language.

[2] Y. Liao, L. Bing, P. Li, S. Shi, W. Lam, T. Zhang, Quase: Sequence editing under quantifiable guidance, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 3855–3864.

[3] I.V. Serban, T. Klinger, G. Tesauro, K. Talamadupula, B. Zhou, Y. Bengio, A. Courville, Multiresolution recurrent neural networks: An application to dialogue response generation, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017.
[4] Y. Hao, H. Liu, S. He, K. Liu, J. Zhao, Pattern-revising enhanced simple question answering over knowledge bases, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 3272–3282.
[5] J. Li, W. Monroe, T. Shi, S. Jean, A. Ritter, D. Jurafsky, Adversarial learning for neural dialogue generation, 2017, arXiv preprint arXiv:1701.06547.
[6] C. Quirk, C. Brockett, W.B. Dolan, Monolingual machine translation for paraphrase generation, in: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 2004, pp. 142–149.
[7] S. Zhao, C. Niu, M. Zhou, T. Liu, S. Li, Combining multiple resources to improve SMT-based paraphrasing model, in: Proceedings of ACL-08: HLT, 2008, pp. 1021–1029.
[8] Y. Gao, W. Zhao, S. Eger, SUPERT: Towards New Frontiers in unsupervised evaluation metrics for multi-document summarization, in: ACL 2020: 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 1347–1354.
[9] I. Sorodoc, K. Gulordava, G. Boleda, Probing for referential information in language models, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 4177–4189.
[10] S. Xu, H. Li, P. Yuan, Y. Wu, X. He, B. Zhou, Self-attention guided copy mechanism for abstractive summarization, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 1355–1362.
[11] X. Zhang, J. Zhong, K. Liu, Wasserstein autoencoders for collaborative filtering, Neural Comput. Appl. (2020) 1–10.
[12] P. Yang, L. Li, F. Luo, T. Liu, X. Sun, Enhancing topic-to-essay generation with external commonsense knowledge, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019.
[13] R. Zhang, C. Chen, Z. Gan, W. Wang, D. Shen, G. Wang, Z. Wen, L. Carin, Improving adversarial text generation by modeling the distant future, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 2516–2531.
[14] A. See, P.J. Liu, C.D. Manning, Get to the point: Summarization with pointer-generator networks, 2017, arXiv preprint arXiv:1704.04368.
[15] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, BLEU: a method for automatic evaluation of machine translation, in: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2002, pp. 311–318.
[16] C.Y. Lin, Rouge: A package for automatic evaluation of summaries, in: Proceedings of the Workshop on Text Summarization Branches Out (was 2004), 2004.
[17] X. Zhang, H. Liu, X. Chen, J. Zhong, D. Wang, A novel hybrid deep recommendation system to differentiate user's preference and item's attractiveness, Inform. Sci. 519 (2020) 306–316.
[18] Y. Zheng, S. Chen, X. Zhang, X. Zhang, X. Yang, D. Wang, Heterogeneous-temporal graph convolutional networks: Make the community detection much better, 2019, arXiv preprint arXiv:1909.10248.
[19] A.D. McCarthy, X. Li, J. Gu, N. Dong, Addressing posterior collapse with mutual information for improved variational neural machine translation, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 8512–8525.
[20] Y. Ren, Z. Wang, Y. Wang, X. Zhang, Generating long financial report using conditional variational autoencoders with knowledge distillation, 2020, arXiv preprint arXiv:2010.12188.
[21] R. Lin, S. Liu, M. Yang, M. Li, M. Zhou, S. Li, Hierarchical recurrent neural network for document modeling, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 899–907.
[22] X. Feng, M. Liu, J. Liu, B. Qin, Y. Sun, T. Liu, Topic-to-essay generation with neural networks., in: IJCAI, 2018, pp. 4078–4084.
[23] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, 2014, arXiv preprint arXiv:1406.1078.
[24] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, 2014, arXiv preprint arXiv:1409.0473.
[25] J. Gu, Z. Lu, H. Li, V.O. Li, Incorporating copying mechanism in sequence-to-sequence learning, 2016, arXiv preprint arXiv:1603.06393.
[26] L. Dong, S. Huang, F. Wei, M. Lapata, M. Zhou, K. Xu, Learning to generate product reviews from attributes, in: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, 2017, pp. 623–632.
[27] N.S. Keskar, B. McCann, L.R. Varshney, C. Xiong, R. Socher, Ctrl: A conditional transformer language model for controllable generation, 2019, arXiv preprint arXiv:1909.05858.
[28] C. Du, H. Sun, J. Wang, Q. Qi, J. Liao, Adversarial and domain-aware BERT for cross-domain sentiment analysis, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 4019–4028.
[29] K. Song, X. Tan, T. Qin, J. Lu, T. Liu, MASS: Masked sequence to sequence pre-training for language generation, in: Proceedings of the Thirty-Sixth International Conference on Machine Learning, 2019.
[30] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R.R. Salakhutdinov, Q.V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, in: Advances in Neural Information Processing Systems, 2019, pp. 5753–5763.
[31] S.R. Bowman, L. Vilnis, O. Vinyals, A.M. Dai, R. Jozefowicz, S. Bengio, Generating sentences from a continuous space, 2015, arXiv preprint arXiv:1511.06349.
[32] Y. Miao, L. Yu, P. Blunsom, Neural variational inference for text processing, in: International Conference on Machine Learning, 2016, pp. 1727–1736.
[33] Y. Miao, P. Blunsom, Language as a latent variable: Discrete generative models for sentence compression, 2016, arXiv preprint arXiv:1609.07317.
[34] S. Semeniuta, A. Severyn, E. Barth, A hybrid convolutional variational autoencoder for text generation, 2017, arXiv preprint arXiv:1702.02390.
[35] P. Yang, L. Li, F. Luo, T. Liu, X. Sun, Enhancing topic-to-essay generation with external commonsense knowledge, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 2002–2012.
[36] T. Wang, X. Wan, T-CVAE: Transformer-based conditioned variational autoencoder for story completion, in: Twenty-Eighth International Joint Conference on Artificial Intelligence IJCAI-19, 2019.
[37] M.H. Yu, J. Li, D. Liu, B. Tang, H. Zhang, D. Zhao, R. Yan, Draft and edit: Automatic storytelling through multi-pass Hierarchical conditional variational autoencoder, AAAI 2020 : Thirty-Fourth AAAI Conf. Artif. Intell. 34 (2) (2020) 1741–1748.
[38] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.
[39] M.J. Kusner, J.M. Hernández-Lobato, Gans for sequences of discrete elements with the gumbel-softmax distribution, 2016, arXiv preprint arXiv:1611.04051.
[40] L. Yu, W. Zhang, J. Wang, Y. Yu, Seqgan: Sequence generative adversarial nets with policy gradient, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017.
[41] K. Lin, D. Li, X. He, Z. Zhang, M.-T. Sun, Adversarial ranking for language generation, in: Advances in Neural Information Processing Systems, 2017, pp. 3155–3165.
[42] W. Fedus, I. Goodfellow, A.M. Dai, MaskGAN: better text generation via filling in the_, 2018, arXiv preprint arXiv:1801.07736.
[43] J. Guo, S. Lu, H. Cai, W. Zhang, Y. Yu, J. Wang, Long text generation via adversarial training with leaked information, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
[44] D. Croce, G. Castellucci, R. Basili, GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples, in: ACL 2020: 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 2114–2119.
[45] C. Xing, W. Wu, Y. Wu, J. Liu, Y. Huang, M. Zhou, W.-Y. Ma, Topic aware neural response generation, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017.
[46] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, 2017, pp. 5998–6008.
[47] M. Freitag, Y. Al-Onaizan, Beam search strategies for neural machine translation, 2017, arXiv preprint arXiv:1702.01806.
[48] I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neural networks, in: Advances in Neural Information Processing Systems, 2014, pp. 3104–3112.
[49] D.P. Kingma, M. Welling, Auto-encoding variational bayes, 2013, arXiv preprint arXiv:1312.6114.