Disentangled Variational Topic Inference for Topic-Accurate Financial Report Generation

Sixing Yan
Department of Computer Science,
Hong Kong Baptist University
Hong Kong SAR, China.
cssxyan@comp.hkbu.edu.hk

Ting Zhu
Research Department, Sales Branch,
TF Securities Co., Ltd.
Shanghai, China.
zhuting@tfzq.com

Abstract

Automatic generating financial report from a set of news is important but challenging. The financial reports is composed of key points of the news and corresponding inferring and reasoning from specialists in financial domain with professional knowledge. The challenges lie in the effective learning of the extra knowledge that is not well presented in the news, and the misalignment between topic of input news and output knowledge in target reports. In this work, we introduce a disentangled variational topic inference approach to learn two latent variables for news and report, respectively. We use a publicly available dataset to evaluate the proposed approach. The results demonstrate its effectiveness of enhancing the language informativeness and the topic accuracy of the generated financial reports.

1 Introduction

Automatically generating long financial reports from a set of macro news have been recently studied with the objective to assist analysts to perform the time-consuming reporting task. A macro news, as shown in Fig. 1, is one paragraph with multiple sentences describing a finance-domain event with supporting details. The corresponding financial report is a longer paragraph with key points of the news and extended analysis, such as inferring and reasoning, with the financial knowledge of analysts. In the literature, long text generation has been well studied in the domain of natural language generation processing (Guo et al., 2018; Guan et al., 2021). Specially, generating long text from the short text with domain-specific settings is still challenging.

The encoder-decoder architecture is commonly employed, where the input news is encoded by a recurrent neural network (RNN) and fed to another RNN model to generate the target report. Some recent works (Beltagy et al., 2020; Chapman et al., 2021) replaced the encoder and decoder with the transformer-based model to learn the long dependency in both news and report text. However, these encoder-decoder models tend to produce generic sentences without the inherent uncertainty in the generated report. This uncertainty arises from the fact that financial reports are written by human specialists with different levels of expertise styles and professional knowledge. Naturally, the reports are very diverse. Probabilistic modeling is reported to be able to learn the uncertainty and diversity of the long texts (Bowman et al., 2016). By learning the stochastic latent variables, the high-level information, such as specialist inference style, is expected to be modeled. Ren et al (Ren et al., 2021b) proposed a variational autoencoder (VAE) method to handle the uncertainty of both news and the report. The background knowledge are learned as the conditional latent variable. In addition, a VAE-based hybrid approach is proposed in (Hu et al., 2020; Ren et al., 2021a) where the report outline is employed as latent variable for VAE decoder. These approaches alleviated the challenges of long text generation. However, the topic of both news and reports are not explicitly learned, where the coherence and coverage of the generated reports are not guaranteed. Recently, in the data-to-report generation domain, Najdenkoska et al (Najdenkoska et al., 2021) proposed a variational model with topic inference to enhance the topic alignment, where a set of latent variables of sentence-level topics are employed. Nevertheless, the topic misalignment between input data and corresponding reports still exists and makes the model hard to be learned.

To address the existing issues of topic modeling and alignment in a unify way, we propose a **Di**s**e**ntangled **V**ariational **T**opic **I**nference (**DeVTI**) approach to generate financial reports by the probabilistic latent variable model. In particular, we learn two disentangled latent variables as the topics of input news and target reports, respectively. The news-related topic represents the context in-

Proceedings of the Fourth Workshop on Financial Technology and Natural Language Processing (FinNLP), pages 18 - 24

News	The European sovereign debt crisis is the manifestation of the "sequelae" of the financial crisis relief policy. The global economy may fall into the stage of "high debt and low growth" due to the sovereign debt crisis or slowdown of the European five countries (PIIGS). The market demand will be weakened, and the process of global recovery from the crisis will be correspondingly prolonged. The window of the Federal Reserve to raise interest rates will be extended to 2011.
Financial Report	The superposition of the external forces of the global economic "rebalancing" and the internal forces of China's economic structural adjustment makes the traditional economic growth mode of China relying on "investment + export" face "passive" adjustment. Under the influence of "real estate regulation + inflation expectation management + European sovereign debt crisis" and other factors, the small cycle of economic downturn has been established; The domestic economic recovery is facing tortuous "Foxconn incident", which will lead to the rise of labor cost and the slow growth of the world economy, which will lead to the decline of China's future economic growth potential. Unlike the economic picture of "two highs and one low" (high unemployment rate, high debt rate and low growth rate) of Western economies, the future picture of China's economy will be: the era of high growth has passed, and it will return from the previous high growth of 11% to the medium growth of 8-10%, The multiple perplexities of "moderate economic growth, moderate structural inflation and low-level overcapacity" are accompanied by controllable inflation: under the background of the establishment of a small cycle of economic downturn, the fall of commodity prices and the lifting of the economic overheating alarm, prices rose in the middle of the year, but inflation is controllable, and the expectation of interest rate increase is weakened. At present, China's macroeconomic policy regulation may be trapped in a "perplexity": China's economy seems to have entered the most contradictory and complex situation, On the one hand, the story of high growth is still expected. On the other hand, the micro operation contradictions highlight the accumulation of many problems, which are almost irreconcilable. In the multi-level goal oriented macro-economic decision-making or the future policy orientation trapped in the macro-economic "maze": (1) the "Chinese version of the national income doubling plan" to stimulate consumption is the key to the switch of economic growth momentum; (2) Economic restructuring: a strategic choice that must be made; (3) The monetary cycle, the economic cycle and the inflation cycle are not synchronized. The economic downturn and policy tightening (liquidity tightening) continued until the end of the third quarter and the beginning of the fourth quarter of 2010. It is expected that the policy will be moderately relaxed at the end of the year; (4) The exchange rate reform was launched, and the interest rate increase was postponed.

Figure 1: An example of news and the corresponding financial report. The co-occurred topics are highlighted.

formation while the report-related topic maintains the prior knowledge of inference and reasoning of human specialists. To summarize, the contributions of this work are three folds,

- We propose a disentangled variational topic inference based approach to generate the topic-accurate financial reports.
- We study the misalignment of the variational topic inference in the short-to-long text generation under the domain-specific setting, and apply disentangled variational inference to learn the latent variables of source and target knowledge individually.
- We demonstrate that our approach achieves comparable performance on a public large-scale news-and-report dataset under a broad range of natural language generation and keyword accuracy evaluation criteria.

2 Methodology

2.1 Preliminaries

Problem Formulation Given the input news X, the goal is to generate a report $Y = \{y_1, y_2, ..\}_{n=1}^{N}$ where y_n refers to the n^{th} sentence. We aim to maximize the conditional log-likelihood as,

$$\theta^* = \arg\max \sum_{n=1}^{N} \log p_\theta(y_n|x), \qquad (1)$$

where θ stands for the model parameters.

Variational Topic Inference To learn to generate report toward the input news, the generation is formulated as a conditional variational inference problem where a set of latent variables z are employed to represent the topics of the report. We incorporate z into the conditional probability of news-based report $\log p_\theta(y|x)$ as,

$$\log p_\theta(y_t|x) = \int \log p_\theta(y_t|x,z) p_\theta(z|x) \mathrm{d}x, \quad (2)$$

where $p_\theta(z|x)$ is the prior distribution condition to the input news x. A variational posterior $q_\phi(z)$ is defined to approximate the intractable true posterior $p_\theta(z|y,x)$ of inferring latent variables z from the input news and the target report. By approximating $D_{\mathrm{KL}}[q_\phi(z)||p_\theta(z|x,y)]$, we can obtain $\mathbb{E}[\log q_\phi(z) - \log p_\theta(y|z,x)p_\theta(z|x)/p_\theta(y|x)] \geq 0$. Following Najdenkoska et al (Najdenkoska et al., 2021), the ELBO of the report generation log-likelihood $\log p_\theta(y|x)$ to be maximized as

$$\log p_\theta(y|x) \geq \mathbb{E}[p_\theta(y|z,x)] - \mathcal{K}_0, \qquad (3)$$

where $\mathcal{K}_0 = D_{\mathrm{KL}}[q_\phi(z|y)||p_\theta(z|x)]$. z is sampled from the variational posterior distribution $z_{train} \sim q_\phi(z_t \mid y)$ in the training, and sampled from the prior distribution $z_{test} \sim p_\theta(z \mid x)$ in the inference.

Misaligned Topic Inference In Eq.(3), the information covered by report y, i.e., $\mathcal{I}(y)$ is assumed to be $\mathcal{I}(y) \subseteq \mathcal{I}(x)$ which is too strong and not hold in practice. The financial news is usually about a particular domain event. However, the financial report is intuitively composed of key points of that event and conclusive inference from business analysts. The logical analysis presented by the financial report is depended on the analyst knowledge and common sense which are not well presented in the input news. Thus, only $\mathcal{I}(y) \cap \mathcal{I}(x) \neq \emptyset$ is guaranteed such that aligning $\mathcal{I}(y)$ and $\mathcal{I}(x)$ by the same latent variable z will incur the misaligned topics.

2.2 Disentangled Variational Topic Inference

The proposed DeVTI model is illustrated in Fig. 2. We first disentangle the topic of news and report from the single latent z by using another VAE to

learn the extra knowledge z_y in the target report y. The corresponding ELBO is given following (Bai et al., 2020),

$$\begin{aligned}\mathcal{O} =& \frac{1}{2}(\mathbb{E}_{z_y\sim q_\psi}[\log p_\theta(y|z_y)] \\ &+ \mathbb{E}_{z_x\sim q_\phi}[\log p_\theta(y|z_x)]) - \mathcal{K}_1,\end{aligned} \tag{4}$$

where $\mathcal{K}_1 = D_{\mathrm{KL}}[q_\psi(z_y|y)||q_\phi(z_x|x)]$. The first term encourages the report reconstruction by z_y while the second term encourages the report generation by z_x. $\mathcal{K}_1$ penalizes the KL divergence between the approximated distribution $q_\psi(z_y|y)$ and $q_\phi(z_x|x)$ which are conditional to y and x. The knowledge l, which is related to news x and extracted from report y, is expected to be aligned as

$$\begin{aligned}\mathcal{K}_2 &= D_{\mathrm{KL}}[p_\theta(l|z_y)||p_\theta(l|z_x)] \\ &= \mathbb{E}_{q_\phi}[\log p_\theta(z_y|l)] - \mathbb{E}_{q_\psi}[\log p_\theta(z_x|l)] - \mathcal{K}_3\end{aligned} \tag{5}$$

where $\mathcal{K}_3 = D_{\mathrm{KL}}[q_\psi(z_x)||q_\phi(z_y)]$. Given that the knowledge is covered by reports as $\mathcal{I}(l) \subset \mathcal{I}(y)$, $\mathbb{E}[\log p(z_y|l)] \leq \mathbb{E}[\log p(z_y|y)]$ is hold,

$$\mathcal{K}_2 \leq \mathbb{E}_{q_\phi}[\log p_\theta(z_y|y)] - \mathbb{E}_{q_\psi}[\log p_\theta(z_x|l)] - \mathcal{K}_3. \tag{6}$$

Thus, a higher lower bound is conducted as,

$$\begin{aligned}\mathcal{O} \leq& \frac{1}{2}(\mathbb{E}_{q_\psi}[\log p_\theta(y|z_y)] \\ &+\mathbb{E}_{q_\phi}[\log p_\theta(y|z_x)]) - \mathcal{K}_4 - \mathcal{K}_2 - \mathcal{K}_3\end{aligned} \tag{7}$$

where $\mathcal{K}_4 = D_{\mathrm{KL}}[p_\theta(z_x|x)||p_\theta(z_x|l)]$. The right-hand-side is the lower bound of objective function Eq.(4) where $\mathcal{K}_4$ penalizes the KL divergence between the approximated distribution $p_\theta(z_x|x)$ and $p_\theta(z_x|l)$. In this way, z_x is disentangled to focus on learning the topic information from input financial news, while z_y is focusing on learning the domain knowledge of target reports. Finally, we are able to learn the model by maximizing a new ELBO as,

$$\begin{aligned}&\frac{1}{2}(\mathbb{E}_{q_\psi}[\log p_\theta(y|z_y)] + \mathbb{E}_{q_\phi}[\log p_\theta(y|z_x)]) \\ &-\beta_2\mathcal{K}_2 - \beta_3\mathcal{K}_3 - \beta_4\mathcal{K}_4\end{aligned} \tag{8}$$

where β_* is the hyper-parameter to control the similarity between several Gaussian distributions (Bai et al., 2020). The $\mathcal{K}_2$ penalizes the KL divergence between the predicted label from generated report and image, which enforces the uncertainty of report to be close to the observed image. The $\mathcal{K}_3$ penalizes the KL divergence between language latent variable $q_\psi(z_y)$ and topic latent variable $q_\phi(z_x)$, which releases the conditions in $\mathcal{K}_1$ and encourages two distribution contain the shared topic knowledge.

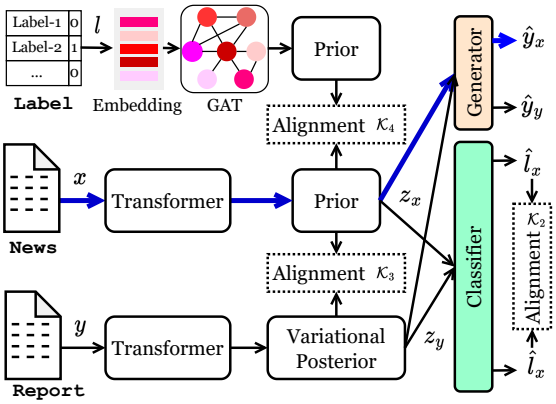

Figure 2: The deep model architecture. In the training, the workflow in black and blue arrow lines are applied while only blue arrow lines are applied in the testing.

	Avg. (Std.)	Percentile 5%	Percentile 95%
# tokens per news	92.6 (±55.5)	58	183
# tokens per report	412.7 (±233.2)	81	784
# sent. per news	1.4 (±1.8)	1	3
# sent. per report	2.1 (±4.4)	1	10
sent. len. per news	66.6 (±41.8)	11	139
sent. len. per report	198.0 (±233.8)	11	635

Table 1: The statistics of the benchmark dataset, including the number of token, sentences, and sentence length for input news and target reports, respectively.

3 Experiments

3.1 Data and Evaluation Criteria

Data We evaluate the proposed approach on the large-scale news-and-report dataset (Ren et al., 2021b). The raw dataset[1] is composed of 10,707 pairs of macro news and corresponding financial reports. We tokenize all news and reports, and filter out frequency least than 5 by an open-source toolkit [2] . This results in 16,052 unique tokens including four special tokens <pad>, <start>, <end> and <unk> (related statistics is shown in Table. 1).

There is no existing topic annotations provided by the raw dataset, so we further automatically annotate each new-and-report pair by the public available tools. We apply a event parser, which is pre-trained on financial knowledge graph data (Wang et al., 2021) , to extract 10 types of entities and 19 types of relationships, and apply a sentiment classifier (Tian et al., 2020) to predict their sentiment polarity (details could be found in A.2). We utilize event subject-predicate-object (SPO) triple

[1] https://github.com/papersharing/news-and-reports-dataset

[2] https://github.com/hankcs/HanLP

Model	NLG Metrics						CA Metrics			
	B.-1	B.-2	B.-3	B.-4	R.	C.	E.	S-E.	ER.	S-ER.
SEQ. (Bahdanau et al., 2014)	32.69	7.65	4.85	2.75	3.59	-	-	-	-	-
SEQA. (Bahdanau et al., 2014)	33.64	13.85	9.89	6.92	3.89	-	-	-	-	-
POINTERNET (See et al., 2017)	36.45	9.51	5.75	2.45	3.44	-	-	-	-	-
CVAE (Zhao et al., 2017)	33.50	14.07	10.04	6.97	4.65	-	-	-	-	-
CVAE-KD (Ren et al., 2021b)	46.67	20.32	12.81	8.00	6.95	-	-	-	-	-
TRANS. (Vaswani et al., 2017)	**48.00**	25.30	9.22	10.65	4.05	39.67	25.10	15.44	9.56	8.03
T-CVAE (Wang and Wan, 2019)	43.01	19.00	13.00	10.98	7.03	34.76	20.33	16.66	13.90	8.47
VTI (Najdenkoska et al., 2021)	40.88	23.43	12.90	10.91	**10.09**	30.65	19.40	17.32	14.89	11.03
DEVTI W/ E.	39.09	**26.70**	12.51	5.57	7.87	33.43	**25.66**	20.30	12.03	10.02
DEVTI W/ S-E.	39.50	22.77	11.32	6.01	6.99	31.32	25.10	**21.30**	12.41	11.02
DEVTI W/ ER.	38.01	20.35	10.32	8.93	6.56	39.03	19.43	17.93	14.90	10.30
DEVTI W/ S-ER. (proposed)	41.70	24.01	**13.90**	**11.11**	9.69	**39.86**	23.59	20.43	**15.09**	**12.33**

Table 2: Performance comparison of report generation models. The experimental results in first section is directly cited from Ren et al (Ren et al., 2021b). The experimental results in second section is our replicated results using their codes. The best scores are in bold face and the second best are underlined."B.", "R." and "C." stand for BLEU, ROUGE and CIDEr scores, respectively. "E.", "S-E.", "ER." and "S-ER." stand for the F-1 measure score of entity, entity with sentimental polarity, entity relationship and entity relationship with sentimental polarity, respectively.

with sentiment polarity to construct labels of the news and report data, respectively.

Evaluation Criteria For report quality, we adopt the natural language generation metrics[3] including BLEU (Papineni et al., 2002), ROUGE-L (Lin, 2004) and CIDEr (Vedantam et al., 2015). To measure the topic accuracy, we adopt the F-1 measure for evaluating the entity and entity relationship with or without sentimental polarity that are extracted from the generated report and ground truth. The micro-avg percentage scores are reported.

3.2 Baseline model and Experiment Setting

We compare the proposed approach with several generation models, including SEQ. (Sutskever et al., 2014), SEQA. (Bahdanau et al., 2014), TRANS. (Vaswani et al., 2017), POINTERNET (See et al., 2017), CVAE (Wang and Wan, 2019), T-CVAE (Wang and Wan, 2019), CVAE-KD (Ren et al., 2021b) and VTI (Najdenkoska et al., 2021). For the proposed DEVTI model, we apply entity relationship with sentimental polarity to optimize the generator, denoted as DEVTI W/ S-ER. The dimensions of hidden state and number of heads in MHA are set as 512 and 8. The model is trained with the learning rate 1e-5 with the mini-batch size of 16. We run the experiments three times with different random seeds and report the average scores. The Implementation details could be found in A.1.

3.3 Experimental Results and Analysis

We evaluate baseline and the proposed approaches by the NLG metrics and classification accuracy metrics in Table 2 1st and 2nd sections.

Performance of Report Generation The proposed DEVTI achieves comparative performances in most of the NLG metrics. In addition, DEVTI achieves best scores in accuracy of entity relationship with sentimental polarity which is more challenging but critical for report usability and reliability. Both results indicate the effectiveness of learning disentangled latent variables for aligning the topics between input news and target reports while ensuring the language informativeness. One possible reason could be that the relationship between entities with sentimental polarity mainly determines the style and topic of the report reasoning and inference. Thus, the latent variable of domain knowledge could enhance the both language fluency and topic accuracy coordinately.

Sensitivity Analysis To analyze how the label affects the report generation performance, we conduct the experiments of learning DEVTI with different labels (as shown in 3rd section). The results consistent with the commonsense that rich semantic knowledge benefit the generation of long and topic-accurate texts with domain-specific setting.

4 Conclusion

In this work, we propose a disentangled variational topic inference (DeVTI) approach to enhance the topic-accurate financial report generation. Two latent variables are learned for the topic of news and extra knowledge of reports. The experiments show the effectiveness of the proposed DeVTI is able to generate descriptive report with correct topics.

[3] https://github.com/tylin/coco-caption

References

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450*.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Junwen Bai, Shufeng Kong, and Carla Gomes. 2020. Disentangled variational autoencoder based multi-label classification with covariance-aware multivariate probit model. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*.

Iz Beltagy, Matthew E Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*.

Samuel R Bowman, Luke Vilnis, Oriol Vinyals, Andrew M Dai, Rafal Jozefowicz, and Samy Bengio. 2016. Generating sentences from a continuous space. In *20th SIGNLL Conference on Computational Natural Language Learning, CoNLL 2016*, pages 10–21. Association for Computational Linguistics (ACL).

Clayton Chapman, Lars Hillebrand, Marc Robin Stenzel, Tobias Deusser, Christian Bauckhage, and Rafet Sifa. 2021. Towards generating financial reports from table data using transformers.

Marcella Cornia, Matteo Stefanini, Lorenzo Baraldi, and Rita Cucchiara. 2020. Meshed-memory transformer for image captioning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10578–10587.

Jian Guan, Xiaoxi Mao, Changjie Fan, Zitao Liu, Wenbiao Ding, and Minlie Huang. 2021. Long text generation by modeling sentence-level and discourse-level coherence. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6379–6393.

Jiaxian Guo, Sidi Lu, Han Cai, Weinan Zhang, Yong Yu, and Jun Wang. 2018. Long text generation via adversarial training with leaked information. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Wenxin Hu, Xiaofeng Zhang, and Yunpeng Ren. 2020. Generating financial reports from macro news via multiple edits neural networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 667–682. Springer.

Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Ivona Najdenkoska, Xiantong Zhen, Marcel Worring, and Ling Shao. 2021. Variational topic inference for chest x-ray report generation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 625–635. Springer.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318.

Yunpeng Ren, Wenxin Hu, Ziao Wang, Xiaofeng Zhang, Yiyuan Wang, and Xuan Wang. 2021a. A hybrid deep generative neural model for financial report generation. *Knowledge-Based Systems*, 227:107093.

Yunpeng Ren, Ziao Wang, Yiyuan Wang, and Xiaofeng Zhang. 2021b. Generating long financial report using conditional variational autoencoders with knowledge distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 15879–15880.

Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27.

Hao Tian, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu. 2020. Skep: Sentiment knowledge enhanced pre-training for sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4067–4076.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Ramakrishna Vedantam, C Lawrence Zitnick, and Devi Parikh. 2015. Cider: Consensus-based image description evaluation. In *Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition*, pages 4566–4575.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Tianming Wang and Xiaojun Wan. 2019. T-cvae: Transformer-based conditioned variational autoencoder for story completion. In *IJCAI*, pages 5233–5239.

Wenguang Wang, Yonglin Xu, Chunhui Du, Yunwen Chen, Yijie Wang, and Hui Wen. 2021. Data set and evaluation of automated construction of financial knowledge graph. *Data Intelligence*, 3(3):418–443.

Di You, Fenglin Liu, Shen Ge, Xiaoxia Xie, Jing Zhang, and Xian Wu. 2021. Aligntransformer: Hierarchical alignment of visual regions and disease tags for medical report generation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 72–82. Springer.

Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi. 2017. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. *arXiv preprint arXiv:1703.10960*.

A Appendix

A.1 Implementation via Deep Neural Network

As common practice in similar research (Kingma and Welling, 2013; Najdenkoska et al., 2021), $q_\phi(z_x|x)$, $q_\phi(z_y|l)$ and $q_\psi(z_y|y)$ are all parameterized as fully factorized Gaussian distributions and inferred by multi-layer perceptrons (MLPs). They are denoted as language prior module, label posterior module and the language posterior module. The proposed DeVTI model is illustrated in Fig. 2. The log-likelihood is implemented as a cross entropy loss based on the generated report and ground-truth reports.

Topic Posterior Modules A matrix E is applied to learn the pre-defined topic embedding. In addition, we also learn the relationship between the topics by the graph attention layer (Veličković et al., 2017). A pair of topics are connected refer to their co-occurrence in the same news-and-report pairs.

Language Prior and Posterior Modules A pre-trained Financial BERT model is employed to learn the token embedding of input text. The input news is fed to the prior module while the target report is fed to the posterior module. Noted that, the posterior module produce the latent topics for guiding the learning the generation, which only applied in the training stage.

Report Generator Module We employ the transformer (Vaswani et al., 2017) as the decoder to generate report. The transformer is composed of multi-head attention module which is able to learn the long dependency in news, report and news-to-report. For each decoding step t, the hidden stats h_t is encoded from the input word features x_t by the standard encoder from Transformer,

$$x_t = w_t + e_t; h_t = \mathrm{MHA}(x_t, x_{1:t}), \quad (9)$$

where w_t and e_t are the word embedding and positional embedding, respectively. A multi-layers transformer decoder is employed to generate the proper report by the latent variable z, following (Cornia et al., 2020; You et al., 2021), h'_t is calculated as,

$$h'_t = \mathrm{MHA}(h_t, z). \quad (10)$$

We apply L-layer MHA where each layer is followed by the operations of residual connection (He et al., 2016) and layer normalization (Ba et al., 2016). Each word is predicted by $y'_t \sim p_t = \mathrm{Softmax}(h'_t W^R)$ where $W^R \in \mathbb{R}^{\mathrm{D}\times\mathrm{W}}$ is the linear projection to transform latent feature into word embedding.

Report Classification Module We employ the fully-connected network with the *Sigmoid* function as the binary classifier to predict each topic from latent variable z,

$$p = \mathrm{Sigmoid}(\max(0, zW_1 + b_1)W_2 + b_2), \quad (11)$$

where $W*$ and $b*$ are learnable parameters. The classifiers are learned by weighted binary cross entropy losses to reduce the label imbalance issue.

A.2 Label Construction

Entity-relationship Extraction We apply a financial research report-based knowledge graph[4] to extract the financial entities and their relationships. The 10 entity types include Industry, Organization, Research report, Risk, Person, Product, Service, Brand, Article and Indicator. The SPO triples of 19 entity relationships include (Industry, subordinate of, Industry), (Organization, belong to, Industry), (Research report, be related to, Industry), (Industry, has, Risk), (Organization, has, Risks), (Organization, be affiliated with, Organization), (Organization, invest, Organization), (Organization, merge, Organization), (Organization, be the customer of, Organization), (Person, work for, Organization), (Person, invest, Organization), (Research report, be related to, Organization), (Organization, produce and sale, Product), (Organization, purchase, Product), (Organization, provide, Service), (Organization, has, Brand), (Product, belong to, Brand),

[4] http://openkg.cn/dataset/fr2kg

(Organization, publish, Article) and (Research report, use, Indicators).

A financial BERT [5] followed a Conditional Random Fields (CRF) model is learned to tag the token with entities and predict the corresponding relationships. The tagging model is trained by the official code[6]. After that, the pre-trained tagging model is applied to extract the entities and their relationships from each sentence of the news-and-report data.

Sentiment Analysis We apply a open-source sentiment analysis toolkit[7] to predict the sentimental polarity of each sentence of the news-and-report data. We set threshold as 0.9 such that one sentence is predicted to be "Positive" or "Negative" only if the related predicted probability is larger than 0.9; otherwise it is predicted to be "Neutral".

The extracted entities and their relationships with the sentimental polarity of each sentence is employed as a label of that sentence, while labels of all sentences are constructed to be the multiple labels of one news or report.

[5] https://github.com/valuesimplex/FinBERT

[6] https://github.com/wgwang/ccks2020-baseline

[7] https://github.com/baidu/Senta