Chinese Reviews Generation Based on HM-BiLSTM Model

Jianglin Yuan¹, Zhigang Guo²*, Gang Chen³, Yihe Sun⁴, RuiPeng Yang⁵

¹ The Information System Engineering Institute, PLA Strategic Support Force Information Engineering University, Zhengzhou, China

*Email: keep_er@163.com

Abstract. HM-BiLSTM model is proposed to solve the problem of inaccurate expression of Chinese comment with specific topic, which is generated by existing deep learning models. Firstly, HM-BiLSTM model was created to generate comments with subject attention mechanism algorithm whose purpose was to assist to constrain the theme. Then, imitation writing sentence was retrieved to extract syntactic structure information from Chinese reviews corpus. The subject as well as structure information was both encoded. Lastly, Chinese comments were generated recurrently by HM-BiLSTM model under the constraints of the subject attention mechanism algorithm and syntactic structure information. The experiment results showed that the model generated comments with better quality and accurate theme.

1. Introduction

There are some achievements in text generation by using neural networks models at present. However, there were few researches on Chinese reviews generation under constraints such as title and sentiment. Xu (2018) used Latent Dirichlet Allocation subject model to generate Hanyu Shuiping Kaoshi essay [1]. It extracted the key sentence to generate the essay under the subject constraints. Liu (2018) used variation probability graph model to generate the text which aimed to satisfy the demands for abstract summary generation [2]. Yizhe Zhang (2017) generated the text by adversarial features matching method whose generator was Long Short-Term Memory Neural Networks and the discriminator was Convolutional Neural Networks [3]. Maosong Sun (2017) generated the poem with RNN Encoder-Decoder model under the subject constraints [4]. Generating poems is easy to realize for the fixed format. But the comments generation is not easy to realize for the flexible format.

McAuley (2012) extracted attributes and perspectives from multidisciplinary reviews [5]. Bing (2016) took an unsupervised approach to extracting attributes from consumer reviews [6]. The extraction of attributes laid foundation for the later generation of product reviews. Dong L (2017) generated commodity comments automatically by using the commodity attributes [7]. However, because the review was limited to the products with fewer attributes, it tended to have grammatical errors.

The research on Chinese comment generation techniques under the subject feature constraints played an important role in improving the quality of generated comments and expanding the application. The method can solve the problem that the current methods generated reviews with inaccurate required subject information. Hence, the Hybrid Multi-Bidirectional Long Short-Term Memory (HM-BiLSTM) model was put forward. Firstly, HM-BiLSTM model was created to generate comments with subject attention mechanism algorithm whose purpose was to assist to constrain the theme. Then, imitation writing sentences were retrieved to extract syntactic structure information from Chinese reviews corpus. The subject as well as structure information was both encoded. Lastly, Chinese comments were generated recurrently by HM-BiLSTM model under the constraints of subject attention mechanism algorithm and
syntactic structure information.

2. Methodology

Fig. 1 illustrated how HM-BiLSTM model generated Chinese reviews.

2.1 The extraction of theme features

The Chinese comments generation under the constraint of subject feature is based on subject prior information. However, different subject prior information has different influence on comments generation. The news’ title was set as parameter $w^{\text{title}} = (w_1^{\text{title}}, ..., w_K^{\text{title}})$. The word embedding was represented as $v^{\text{title}} = (v_1^{\text{title}}, v_2^{\text{title}}, ..., v_K^{\text{title}})$.

The syntactic constraints were extracted to generate reviews from the comment corpus. The imitation writing sentence was represented as parameter $w' = (w_1', ..., w_L')$. At the same time, the word embedding was represented as parameter $v' = (v_1', ..., v_L')$. After part-of-speech tagging [8][9], the key structural information were retained as the structural characteristics constraints for comments generation.

2.2 HM-BiLSTM Model

Since the language model was more dependent on the historical and future information for generating current word, the HM-BiLSTM model was proposed in this paper. The model improved the interaction between historical and future information, which were used to predict the vocabulary. There were three bidirectional Long Short-Term Memory neural networks combined as HM-BiLSTM model. The model adopted the skipping lexical prediction mechanism and integrating multiple BiLSTM models which enriched historical and future information to increase the accuracy of lexical prediction. The formulation was shown below:

$$h_{i+1} = f(\text{concat}(h^1_{i+1}, h^2_{i+1}, h^3_{i+1}) \cdot W_o + b_o) \quad (1)$$

$h^1_{i+1}$, $h^2_{i+1}$ and $h^3_{i+1}$ represented the hidden output of BiLSTM1, BiLSTM2 and BiLSTM3, respectively. Variable $h_{i+1}$ represented the result of the combination of multi-hidden layer states. Function concat(·) represented the hidden layer output vector splicing function. Function $f(·)$ represented nonlinear function. Variable $W_o$ represented parameter matrix and Variable $b_o$ represented bias matrix which were both needed to be trained in neural networks. The formulations were shown below:

$$h^1_i = \text{BiLSTM1}(y_{i-1}) \quad (2)$$
$$h^2_{i+1} = \text{BiLSTM2}(y_i) \quad (3)$$
$$h^3_{i+1} = \text{BiLSTM3}(y_{i+2}) \quad (4)$$
Variable $y'_{i+1}$ represented the result of output layer at time $i+1$. The formulation was shown below:

$$ y'_{i+1} = g(h_{i+1} \cdot W + b) $$

(5)

Variable $W$ and Variable $b$ represented parameters matrix. Function $g()$ was nonlinear function. The output of the model was calculated with following formulation below:

$$ y_{i+1} = \text{title}_\text{att}(y'_{i+1}, S', \text{title}, p, T) $$

(6)

Variable $S'$ and Variable $\text{title}$ represented the structure of sentence and subject features vectors. According to subject attention mechanism, HM-BiLSTM model generated reviews under subject features constraints by combining syntactic structure characteristic constraints $T$ and subject similarity constraint threshold $p$. Parameters $Y = \{y_1, y_2, ..., y_n\}$ represented the generated vocabulary set. Function $\text{title}_\text{att}()$ represented the subject attention mechanism algorithm which would be elaborated.

### 2.3 Subject Attention Mechanism Algorithm

Subject attention mechanism algorithm was mainly used to constrain the theme of sentence when generating text according to the structure information and subject constraints. There were lots of research on Chinese particles and part-of-speech tagging which have some mature algorithms [10][11]. The process of subject attention mechanism algorithm is shown in Table 1.

**Table 1. Subject attention mechanism algorithm**

| Subject Attention Mechanism Algorithm |
|---------------------------------------|
| **1** Random initialize HM-BiLSTM model; |
| **2** for $i <= n$ : |
| **3** Input $S'$ for calculating $y'_{i+1}$; |
| **4** The first $m$ vocabs were selected according to the probability distribution of $y'_{i+1}$ under the constraints $t_i$. The selected vocabulary was $V = \{v_1, v_2, ..., v_m\}$ under the constraints: |
| **5** The word is added to $S'_i = \{y_1, ..., y_i, v_j, s'_{i+2}, ..., s'_{a}\}$ $v_j \in V$, $S'' = \{S'_1, ..., S'_a\}$: |
| **6** The subject similarity between reviews and news were calculated with cosin functions: |
| **7** The sentence is selected from $S''$ when the similarity of title is greater than $p$. The word $v_j$ is extracted as $y_{i+1}$. Updating $S'_i = \{y_1, ..., y_i, v_j, s'_{i+2}, ..., s'_{a}\}$ |
| **8** END |

### 3. Experiment Settings

#### 3.1 Data and Processing

There were some news and comments collected from website to be the training dataset and test dataset for text generation [12]. The detail dataset was shown in Table 2.

**Table 2. The news dataset**

| Type      | Training dataset | Test dataset |
|-----------|------------------|--------------|
| news (item) | 445000 | 50000 |
| comments (item) | 660000 | 750000 |

Preprocessing of data mainly carried on the word segmentation and the part of speech annotation to the selected comment sentence. The news demands for expanding huge vocabulary to improve the accuracy of word segmentation and the part of speech annotation, which has great significance for model training.
3.2 Model Training and Evaluation
The dimension of word embedding was set to 300. The hidden layer neurons units were 1500. The layer of networks was set to 5. In order to prevent overfitting in model training, the dropout layer was adopted whose coefficient was set to 0.5. The loss function was cross-entropy and the error transferred by using the AdaDelta algorithm when training the model [13][14].

The topic similarity threshold \( p \) was set in subject attention mechanism algorithm. Different topic similarity thresholds had different influences on the quality of generated comments. The model training process was completed when the model converged to an error less than 0.5.

The evaluation is very important for text generation which includes subjective evaluation metrics and objective evaluation metrics. The subjective evaluation metrics are mainly dependent on judgment by experts. The objective evaluation metrics are mainly dependent on the corpus and algorithm. The BLEU (Bilingual Evaluation Understudy) algorithm and ROUGE (Recall-Oriented Understudy for Gisting Evaluation) algorithm are the main method to evaluate text.

3.3 Results and Analysis

3.3.1 The influence of topic similarity threshold on comment quality.
Table 3 was the result which evaluated the quality of generated comments with different subject similarity thresholds. It measured the quality of the generated comments with subjective evaluation metrics.

| \( p \) | Fluency | Readability | Novelty |
|-------|---------|-------------|---------|
| 0.50  | 3.6     | 3.7         | 3.8     |
| 0.55  | 3.8     | 3.9         | 4.0     |
| 0.60  | 4.0     | 4.1         | 3.6     |
| 0.65  | 4.3     | 4.2         | 3.9     |
| 0.70  | 4.5     | 4.2         | 4.1     |
| 0.75  | 4.1     | 3.9         | 4.0     |
| 0.80  | 3.2     | 3.5         | 3.2     |
| 0.85  | 2.9     | 3.2         | 2.5     |

As shown in Table 3, when the topic similarity threshold was set to around 0.70, the comments qualities were good. When the topic similarity threshold was too low or too high, the qualities of the generated comments were bad. When the threshold value was too low, it increased the uncertainty for the vocabulary prediction. When the threshold value was too high, the model was easy to overfitting. The topic interfered too much with the lexical prediction, resulting in the deterioration of the text generated by the model.

3.3.2 Objective evaluation result
Objective evaluation criteria evaluated the generated comments qualities by BLEU algorithm and ROUGE algorithm. The results were shown Table 4.

| MODELS     | BLEU-2 | BLEU-3 | BLEU-4 | ROUGE |
|------------|--------|--------|--------|-------|
| Seq2Seq    | 0.35   | 0.34   | 0.23   | 0.36  |
| VAE        | 0.30   | 0.29   | 0.22   | 0.29  |
| SeqGAN     | 0.38   | 0.33   | 0.32   | 0.37  |
| HM-BiLSTM  | 0.43   | 0.39   | 0.31   | 0.45  |
According to analysis from Table4, the qualities of comments generated by HM-BiLSTM model were higher in BLEU-2, BLEU-3, BLEU-4 and ROUGE scores than the other models. But the BLEU-4 scores were slightly lower than SeqGAN model.

Through comprehensive analysis, HM-BiLSTM model based on thematic attention mechanism made better use of historical and future information to predict lexical more accurately. By using the syntactic structure information of the commentss sentences imitated, the model combined the news and comments prior information to generate comments more relevant to the news topic.

3.3.3 Evaluation results of topic relevance
The subject similarity threshold was set to 0.70. The evaluation results were shown in Table 5. The thematic attention mechanism is used in HM-BiLSTM model, which has a great improvement in thematic relevance.

| Models   | accuracy | error  |
|----------|----------|--------|
| Seq2Seq  | 0.740    | 0.260  |
| VAE      | 0.605    | 0.395  |
| SeqGAN   | 0.825    | 0.175  |
| HM-BiLSTM| 0.905    | 0.095  |

4. Conclusions
The HM-BiLSTM model can solve the problem that the current methods generate comments with inaccurate required subject information. Since the syntactic structure is used as the structural constraint of HM-BiLSTM model, the model reduces the dependence on the automatic learning of grammatical structure. The subject attention mechanism is used as the subject constraint of HM-BiLSTM model to reduce the difficulty and complexity of generating comments. Therefore, it reduces syntax errors and generates comments with more relevant subject. The experiment results show that the quality of comments which are generated by HM-BiLSTM model is higher in subjective and objective evaluation scores, subject relevance when subject similarity threshold is set at 0.7.

Reference
[1] Yanhua Xu,Yujie Miao,Lin Miao,Xueqiang Lv. Generating HSK Writing Essays with LDA Model[J]. Data Analysis and Knowledge Discovery, 2018, 2(9): 80-87.
[2] Liu Tingzhen, Zhang Hua. Text generation algorithm based on variant probability graph model [J]. Computer application, 2018 (2018 z1): 99-103.
[3] Zhang Y, Gan Z, Fan K, et al. Adversarial feature matching for text generation[C]//Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017: 4006-4015.
[4] Yi X, Li R, Sun M. Generating chinese classical poems with rnn encoder-decoder[M]/Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Springer, Cham, 2017: 211-223.
[5] McAuley J, Leskovec J, Jurafsky D. Learning attitudes and attributes from multi-aspect reviews[C]//2012 IEEE 12th International Conference on Data Mining. IEEE, 2012: 1020-1025.
[6] Bing L, Wong T L, Lam W. Unsupervised Extraction of Popular Product Attributes from E-Commerce Web Sites by Considering Customer Reviews[J]. Acm Transactions on Internet Technology, 2016, 16(2):12.
[7] Dong L, Huang S, Wei F, et al. Learning to generate product reviews from attributes[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. 2017, 1: 623-632.
[8] Liang X T, Gu L. Research on Chinese Word Segmentation and Part-Of-Speech Tagging [J]. Computer Technology and Development, 2015 (02): 175-180.
[9] Si N W, Wang H J, Li W. Chinese Part-of-Speech Labeling Model Based on Attention-span Memory Network [J]. Computer Science, 2018, 4:009.

[10] Ding J, Zhao J H. Research on Chinese Word Segmentation Algorithm Based on n-gram Model [J]. Fujian Computer, 2017, 33(5): 110-110.

[11] Hu J, Zhang J C. Chinese Word Segmentation Model of Bidirectional Cyclic Network [J]. Minicomputer System, 2017, 38(3):522-526.

[12] Lianhui Qin, Lemao Liu, Wei Bi, Yan Wang, Xiaojiang Liu, Zhiting Hu, Hai Zhao, and Shuming Shi, “Automatic article commenting: the task and dataset,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, (Volume 2: Short Papers). 2018, pp. 151–156.

[13] De Boer P T, Kroese D P, Mannor S, et al. A tutorial on the cross-entropy method[J]. Annals of operations research, 2005, 134(1): 19-67.

[14] Zeiler M. ADADELTA: An adaptive learning rate method. arXiv preprint arXiv[J]. Volume, 2012, 1212: 5701.

[15] Papineni K, Roukos S, Ward T, et al. BLEU: a method for automatic evaluation of machine translation[C]// Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002:311-318.

[16] Arthur P, Neubig G, Nakamura S. Incorporating Discrete Translation Lexicons into Neural Machine Translation[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016: 1557-1567.

[17] Zhang J, Wang X L, Xu H B. Summary of automatic abstract evaluation methods[J]. Chinese journal of information technology, 2008, 22(3):81-88.

[18] Zhang W N, Zhang Y Z, Liu T. Review of systematic evaluation methods of dialogues[J] China science: Chinese science: information science, 2017, 43(8): 954-966.