Sememe-based Topic-to-Essay Generation with Neural Networks

Dan Luo¹, Xinyi Ning² and Chunhua Wu¹*

¹Department of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing, 100876, China
²Department of International, Beijing University of Posts and Telecommunications, Beijing, 100876, China
*Email: wuchunhua@bupt.edu.cn

Abstract. Topic-to-Essay generation task aims to generate topic related and coherence text based on user input. Previous research generates text solely based on the input words and the generation results in not satisfactory. The traditional methods using semantic relationships of words from corpus to guide the model, which made the model rely heavily on the corpus. In this paper, we propose a sememe-based topic-to-essay generation model(S-TEG), which integrates sememes from external knowledge graph HowNet with input topic words to guide the model. In order to prevent introducing noise, we elaborately devise measuring the similarity of the non-current topic words to filter sememes information. The experiment results demonstrate that our approach achieved 4.10 average score in subjective evaluation and a 3.60 BLEU score, which shows that our model is able to generate text that is more coherence, topic-related and fits the daily logic.

1. Introduction
The Topic-to-essay generation (TEG) task is a branch of natural language generation (NLG). It takes user-defined keywords as input and output a paragraph. The aim is to generate a paragraph that is both topic-related and coherent [1]. Even NLG has become a promising area in machine learning, there is little research on the TEG task. Feng et al. first formally proposed the TEG task in 2018 [1]. However, the result is not satisfactory. The generation performance can be random and lack of relevance to the topic. The basic reason that the input information is too insufficient to guide the model. Corpus and knowledge graph are two commonly used as information sources in NLG. Using hidden semantic information in the corpus to guide the model is a traditional approach but it makes the model rely on the corpus quality. What’s more, people write texts based on the known commonsense information, which will not be mentioned in the text. Corpus-based models lack daily knowledge and are difficult to apply to other fields. External knowledge graph has also been adopted widely to provide additional information. A large number of real-world entities and their relationships between them are stored in the semantic knowledge graph. Using the commonsense knowledge graph to expand the input topic information can help the model obtain information that conforms to daily logic.

In this paper, to guide the model with correct semantic information, we propose a novel sememes-augmented neural network for TEG, Sememes-based Topic-to-essay Generation (STEG). In order to prevent introducing noise, we utilize the semantic information of the other topic words to filter the sememes, which can also help to solve the polysemy problem in text generation.
2. Related Work
There is no dedicated research on the TEG task before Feng et al. [1]. Poetry generation is a similar task as it also generating text based on topics. It is important to include themes in the generative poetry model [2]. He et al. assumes that the main topic of a poem can be expressed by users through a few keywords. A set of keywords given by users is accepted as the main topic [3]. Zhang and Lapata introduced a jointly model which also created the first line based on the keywords and generated the subsequent lines based on the generated lines [4]. Wang et al. proposed a model that generates sub-topics based on the user’s input to guide the poetry generation [5]. Ghazvininejad et al. proposed Hafez for generating topical poetry, which will build a scored list related to user-supplied topics [6]. The difference is that poem generation has clearer structure regulations and it is ok to transmit a blurry message while TEG does not have clear structure rules and should convey a clear message [7]. TEG task has been treated as sequence-to-sequence problem. Feng et al. developed the MTA-LSTM model which is the first research on the TEG task. The MTA-LSTM model is build based on a traditional encoder-decoder model. It modifies the traditional attention model by introducing the coverage mechanism to build the coverage vector to enhance the multi-topic awareness [1]. However, the generation quality is not satisfactory. The same word or sentence can be repeated and lack novelty as well.

HowNet is a bilingual (English-Chinese) commonsense knowledge base that provides both inter-conceptual relations and inter-attribute relations of concepts in lexicons [8]. After more than 30 years of hard-work, HowNet now has marked more than 100,000 common words and the total number of the dictionary exceeds 120,000. HowNet committed to model the relationships between concepts and attributes. In HowNet, concepts are semantic descriptions of words and each word can be expressed by several sememes. Sememes are the smallest units of meaning. Through a hierarchical organization of sememes rather than simply classification or multi-layer classification, HowNet reveals the connections between concepts and words. HowNet has been widely used as a knowledge source, for example, machine translation [9], dialogue generation [10].

3. Proposed Method

3.1. HowNet and ConceptNet
Some knowledge graphs, such as ConceptNet, using concepts as basic units to describe the meaning of words. Every concept is a node in the hierarchy. However, sememes are used as nodes to form the hierarchy structure in HowNet. Figure 1 intuitively shows that how “Apple” is expanded in the HowNet and ConceptNet. We simply translate the Chinese results in this paper.

Figure 1(a). HowNet structure example. Figure 1(b). ConceptNet structure example
Typically, we use the top sememe word as the sense for HowNet using a hierarchy structure and the leading sememe shows the main semantic information. There mainly two differences between HowNet and ConceptNet. First, ConceptNet expands “Apple” with adjacent concepts while HowNet defines and expands the basic meaning. Second, “Apple” mainly has two senses in ConceptNet, fruit and tree, but has four senses in HowNet, special brand, fruit, tree, and tool. HowNet sememes give more advanced and clearer semantic background information that is closer to what we want.

3.2. Sememe-based Topic-to-Essay Generation Model

Given a sequence $m$ labeled with $X$ topics, $X$ is the topic word set that contains $n$ topic words $X = \{topic_1, topic_2, topic_3, ..., topic_n\}$. Our goal is to generate a topic related sequence $y$. First, we illustrate the process of obtaining and selecting the sememes in HowNet. Then we discuss how the sememes are used to train the model. Figure 2 shows the framework of the proposed model.

![Figure 2](image)

**Figure 2.** The sketch of our proposed S-TEG model.

In practice, we search each topic word $x_i$ in the HowNet and the retrieved results are used as the optional topic information. The results are stored as different sets according to the leading sememe words. Especially, for the current topic word $x_i$ the searching result $r_i$ is:

$$s_i = \{ \{sense_j, sememe_1, sememe_2, sememe_3, ..., sememe_L \} \}_{j=1}^{K}$$

where $K$ is the number of retrieved senses and $L$ is the number of retrieved sememes sets, $sememe$ is the embedding of sememe words, $sense_j$ is the top sememe which is the sense shows in Figure 1(a) and $[sememe]$ means the concatenation of all the sememes’ embedding. After obtaining the optional appending sets, we precisely select the expanding sememes by measuring the similarity with the topic semantic. We use an average vector $c_i$ to represent the topic semantic. The average vector is calculated based on all the other topics embedding except the current topic word as:

$$c_i = \frac{\sum_{j=1}^{N} (x_j \neq i)}{N}$$

where $N$ is the number of all the topics, $i$ is the subscript of the current word and $x_j$ is the embedding of topic words’. The average vector is used as a representation of other topic words to further guide the selection of sememes. The correlation must between the different sense words and the average vector must be measured. We simply calculate the cosine similarity of them:

$$Similarity_i = \cos (r_i, c_i) = \left\{ \frac{\text{sense}_j \cdot c_i}{\| \text{sense}_j \| \cdot \| c_i \|} \right\}_{j=1}^{K}$$

the higher the similarity score, the closer to the average vector and the closer to other topic words. Through sorting the similarity scores as the following equation, we finally select the additional topic sememe set $r'_j$:

$$n_j = \left( n_j \mid \max(Similarity_i) \right), \quad j = 1, 2, 3, ..., K$$

(4)
then the final input topic is obtained through simple concatenation of the embeddings of sememe set \( r_f \) and current topic word \( x_i \).

\[
\text{Topic} = \{\text{topic}_1, r_1, \text{topic}_2, r_2, \text{topic}_3, r_3 \ldots \text{topic}_n, r_n\} \quad (5)
\]

we use LSTM (42) to calculate the representation \( T_t \) of final input topic word as:

\[
T_t = \sum_{i=1}^{N} \alpha_{t,i} \text{Topic}_i \quad (6)
\]

\[
\alpha_{t,i} = \frac{\exp(g_{t,i})}{\sum_{j=1}^{N} \exp(g_{t,j})} \quad (7)
\]

where \( \alpha_{t,j} \) is the attention weight of the final input topic words. To fully express every topic, the attention mechanism is adopted followed [1] to dynamically control the attention weight as:

\[
z_{t,i} = z_{t-1,i} \frac{1}{\Phi_i} \alpha_{t,i} \quad (8)
\]

\[
g_{ij} = z_{i-1,j} v_{\alpha}^T \tanh(W_{\alpha} h_{t-1} + U_{\alpha} \text{topic}_j) \quad (9)
\]

where \( \Phi_i \) is the parameter used to update the attention vector \( z_t \), \( g_{ij} \) is the attention score within which \( v_{\alpha}, W_{\alpha}, U_{\alpha} \) are three matrices that need to be optimized in the training phase. The probability of the next word is defined as:

\[
P(y_t | y_{t-1}, T_t, z_{t}) = \text{softmax}(g(h_t)) \quad (10)
\]

and the hidden state is updated before each prediction by:

\[
h_t = f(h_{t-1}, y_{t-1}, T_t) \quad (11)
\]

where \( g(\cdot) \) is a linear function and \( f(\cdot) \) is an activation function that is defined by the LSTM structure.

In the training phase, we take the end-to-end learning for the sememe-based TEG model following Feng et al. [1], which learns both the parameters for the generation model (i.e., \( \theta \) for decoding LSTM) and the parameters for guidance (i.e., \( \eta \)). Then the likelihood of reference of the training corpus is maximized through training all the parameters:

\[
(\theta^*, \eta^*) = \arg \max_{\theta, \eta} \sum_{t=1}^{N} \log P(y_t | T_t; \theta, \eta) \quad (12)
\]

4. Experiments and Results

4.1. Datasets and Settings

We conduct experiments on the Essay dataset which is built in the prior research [1]. It consists of 3.2 million Chinese junior or high school essays with which the length is between 50 to 100. Each of them is labeled with five topics. Table 1 shows how we utilize the dataset. We basic follow Feng et al. [1] to set the parameters of the model.

| Essay | Training | Testing |
|-------|----------|---------|
| 4,94,944 | 300,000 | 5,000 |

4.2. Baselines

We compare our proposed model with the following methods.

- **PNN**: A topic-consistent poetry generation method using neural network based on planning [11].
- **MTA-LSTM**: This model utilizes the attention mechanism for the TEG[1].
- **Corpus-TEG**: Using the corpus itself as the source for additional topic information.
- **Concept-TEG**: Using the ConceptNet as the source for additional topic information with the filtering mechanism.
- **HowNet-TEG**: Using the HowNet as the source for additional topic information but appending the additional information without filtering mechanism.
4.3. Evaluation Metrics

We following the previous work [1] adopt two evaluation methods:

**BLEU Evaluation.** Bilingual Evaluation Understudy (BLEU) [12] is an automatic evaluation metric for machine translation. We adopt the BLEU-2 score which evaluates the n-gram overlap for essay generation. The higher the scores the better the results.

**Human Evaluation.** Five Chinese experts were invited to measure the quality of generated essays on four perspectives: “Integrity”, “Relevance”, “Coherence”, “Fluency”. 200 items are selected randomly selected with corresponding topic labels. The experts score them ranged from 1 to 5.

4.4. Results and Discussion

Table 2 shows the human evaluation and BLEU evaluation results. Typically using knowledge graphs to expand topic information has better results than using the corpus. Using HowNet sememes gain higher scores than using ConceptNet. S-TEG using HowNet sememes considering other topic words semantic information gains the highest score. S-TEG improves the integrity and coherence of the generated text. The BLEU evaluation results which are consist of human evaluation results. Our proposed method gains the highest scores on both evaluation metrics.

**Table 2. Evaluation results.**

| Model           | Human Evaluation Score | BLEU Score |
|-----------------|------------------------|------------|
|                 | Integrity | Relevance | Fluency | Coherence | Average Score |
| PNN             | 2.46      | 2.77      | 3.67    | 2.25      | 2.79          | 1.39        |
| MTA-LSTM        | 3.40      | 2.80      | 3.72    | 2.84      | 3.19          | 3.19        |
| C-TEG           | 3.43      | 3.37      | 3.10    | 3.05      | 3.24          | 3.36        |
| ConceptNet-TEG  | 3.50      | 3.32      | 3.38    | 3.35      | 3.39          | 3.47        |
| HowNet-TEG      | 3.63      | 3.62      | 3.63    | 3.43      | 3.58          | 3.56        |
| S-TEG           | 4.50      | 3.88      | 3.87    | 4.13      | 4.10          | 3.60        |

4.5. Case Study

**Table 3. Sememes selection results samples.**

| Topic       | Selected Sememes | Topic       | Selected Sememes |
|-------------|------------------|-------------|------------------|
| Apple 苹果 | Computer 电脑   | Apple 苹果 | Fruit 水果       |
| Publish 发布| Release 发表   | Regimen 养生 | Regimen 养生   |
| Company 公司 | Company 公司 | Diet 饮食 | Food 食物       |
| Modern 现代 | Present 现在 | Life 生活 | Living 活着    |
| Technology 科技 | Knowledge 知识 | Modern 现代 | Present 现在   |

**Table 4. The generation results samples.**

| Topic       | Sample Text                                                                 |
|-------------|----------------------------------------------------------------------------|
| MTA-LSTM    | My dream is to be a scientist and to invent a scientific science. I think that the future of the future will surely be realized. |
| C-TEG       | My dream is to be a scientist and to invent a new kind of science to make our lives better and to make our motherland more prosperous and more powerful! |
| S-TEG       | I could not help to say that how advanced is the future of science and technology! This sentence is current. Now I have a dream. I must study hard and I will study hard to serve my motherland when I grow up. |
Table 3 shows two examples of using the proposed method to append topic words with sememes. The left column is the giving topic words and the right column list corresponding selected additional sememe information. “Apple” is expanded with different specific meanings when companied by different topic words, which is close with human logic.

Table 4 lists three samples that show the generation performance with no additional topic information (MTA-LSTM), using co-occurrence relationships of words from the corpus (C-TEG) and our proposed method (S-TEG) expanding topic information with denoising sememes. Through C-TEG has improved the generation text quality, both C-TEG and MTA-LSTM shows a poor understanding of the “Science” topic. S-TEG using the sememes to define the topic words provides the model more common-sense information.

5. Conclusions
In this paper, we propose using a sememe-based neural model (S-TEG) to generate essays based on users’ input. The sememes information is obtained by using the HowNet knowledge graph. To filter the sememe information, we measure the similarity between the average vector of other topic words and the sememes. This is the first time that other topic semantic information is used in expanding additional information of the current topic word. Through our mode has improved the generation performance, there may still be repeated phenomena. In the future, we plan to change and improve the model structure or beam search strategy to solve this problem.

References
[1] Feng, X., Liu, M., Liu, J., Qin, B., Sun, Y., & Liu, T. (2018, July). Topic-to-Essay Generation with Neural Networks. In IJCAI (pp. 4078-4084).
[2] Hopkins, J., & Kiela, D. (2017, July). Automatically generating rhythmic verse with neural networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 168-178).
[3] He, J., Zhou, M., & Jiang, L. (2012, July). Generating chinese classical poems with statistical machine translation models. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 26, No. 1).
[4] Zhang, X., & Lapata, M. (2014, October). Chinese poetry generation with recurrent neural networks. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 670-680).
[5] Wang, Z., He, W., Wu, H., Wu, H., Li, W., Wang, H., & Chen, E. (2016). Chinese poetry generation with planning based neural network. arXiv preprint arXiv:1610.09889.
[6] Ghazvininejad, M., Shi, X., Choi, Y., & Knight, K. (2016, November). Generating topical poetry. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (pp. 1183-1191).
[7] Oliveira, H. G. (2017, September). A survey on intelligent poetry generation: Languages, features, techniques, reutilisation and evaluation. In Proceedings of the 10th International Conference on Natural Language Generation (pp. 11-20).
[8] Cao, J., Wang, T., Li, W., & Tao, C. (2019, July). A Method of Calculating the Semantic Similarity Between English and Chinese Concepts. In International Conference on Machine Learning and Intelligent Communications (pp. 313-324). Springer, Cham.
[9] Liu, S., Xu, J., & Ren, X. (2019, October). Evaluating semantic rationality of a sentence: A sememe-word-matching neural network based on hownet. In CCF International Conference on Natural Language Processing and Chinese Computing (pp. 787-800). Springer, Cham.
[10] Ma, Z., Du, B., Shen, J., Yang, R., & Wan, J. (2020). An Encoding Mechanism for Seq2Seq based Multi-Turn Sentimental Dialogue Generation Model. Procedia Computer Science, 174, 412-418.
[11] Wang, Z., He, W., Wu, H., Wu, H., Li, W., Wang, H., & Chen, E. (2016). Chinese poetry generation with planning based neural network. arXiv preprint arXiv:1610.09889.

[12] Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).