VV-Couplet: An open source Chinese couplet generation system

Jiyuan Zhang[†‡], Zheling Zhang[†‡] and Shiyue Zhang[†‡] and Dong Wang[†‡*]
[†] Center for Speech and Language Technologies, Research Institute of Information Technology
Department of Computer Science and Technology, Tsinghua University, China
[‡] Beijing National Research Center for Information Science and Technology

***Abstract*—Chinese couplet is a special form of poetry involving a pair of lines that adhere to certain rules. This paper presents VV-Couplet, an open source automatic couplet generation system. This system is based on the attention-based sequence-to-sequence neural model that 'translates' the first line to the second line. Particularly, it relies on a local attention mechanism that learns the basic rules of couplet generation. Moreover, entity names such as person names and addresses are particularly treated. We open an online couplet generation service, and the entire source code and data for interested readers to reproduce our work.**

I. Introduction

Chinese Couplet is a special form of literature, and is often regarded as a special, compact form of poetry. A couplet consists of two lines (antecedent line and subsequent line, respectively) that are in equal length and adheres to some special rules of 'coupling'. Couplets are mostly used in special events (e.g., wedding, birthday) or holidays (e.g., New Year, National Day) to express particular emotions, such as congratulation, condolence, encouraging, etc. Couplets also appear in other forms of poetry to improve tonal aesthesis. Chinese couplet has a long history back to Five Dynasties (10th century A.D.), and has been popular for more than 1,000 years till today. Fig. 1 shows a couplet on the sides of a door, for celebrating the Chinese New Year.

Fig. 1. A couplet on the sides of a door, for celebrating the Chinese New Year. This picture is from baidu.com.

The basic rules of couplets are 'inverse in tone, identical in lexical category'. Besides of these basic rules, each line should be itself meaningful and coherent, and at the same time the meaning and aesthetic sense of the two lines should be closely related. An example is shown in Fig. 2. In this example, the tones of the characters at the same position in two lines are exactly inverse, and their lexical categories (noun, verb, etc.) are the same. Moreover, the meaning and aesthetic sense of the two lines are highly correlated: both describes a pastoral life. More complicated couplets involve special constraints among the characters, e.g., the couplet shown in Fig. 3, where the name of a famous compere appears at the beginning and a key concept repeats several times in each line.

空 山 雨 后 拾 新 翠
(P P Z Z P P Z)
古 寺 云 中 觅 旧 踪
(Z Z P P Z Z P)

Fig. 2. An example of Chinese couplets. The first line is the antecedent line, and the meaning is 'enjoying green mountain after raining'; the second line is the subsequent line, with the meaning 'looking for heritage in the ancient temple'. 'P' and 'Z' represent two tones respectively: Ping and Ze.

董卿观花花开花落花容晚
曹颖望月月隐月圆月色浓

Fig. 3. An example of Chinese couplets with additional constrains, where the name of a famous compere appears at the beginning and a key concept repeats several times in each line.

Due to the strict constraints in tonal patterns and semantic meanings, couplet generation, i.e., generating the subsequent line given an antecedent line, is regarded as highly challenging, and only very few educated people can complete this task. In ancient China, couplets are often used to quiz the capability of intellectuals, in both knowledge and intelligence. In modern China, there are few people can produce good couplets, and most of modern couplets do not strictly comply with the basic rules. Automatic generation of couplets is therefore highly attractive. The immediate merit of automatic couplet generation is in entertainment and education, but the deeper value is to help protect this invaluable cultural heritage.

Research on couplet generation is not extensive. Existing approaches can be categorized into three groups: probabilistic

978-988-14768-5-2 ©2018 APSIPA

model approach, statistical machine translation (SMT) approach, and neural approach. All of these approaches establish a statistically conditional model that describes the generation process conditioned on the given antecedent line.

The probabilistic model approach establishes a probability model that learns the probability of generating each word given the antecedent line. For example, Yi et al. [1] constructed a hidden Markov model (HMM) and treated characters in the subsequent line as hidden variables. Zhang et al. [2] proposed a maximum entropy Markov model (MEMM) that generates the subsequent line character by character, where the generation of each character depends on the preceding generation as well as the characters at or near the same position in the antecedent line.

The SMT approach takes the couplet generation task as a SMT task, where the antecedent and the subsequent lines are regarded as two languages. Zhou et al. [3], [4], [5] carefully studied this approach and published their online couplet generation system[1]. As far as we concerned, this is the only public couplet generation service so far.

Recently, deep neural networks (DNN) have gained significant success on numerous tasks including natural language generation (NLG) [6], [7]. One of the most important DNN models used in present NLG tasks is the sequence-to-sequence model [8], particularly with the attention mechanism [9]. This model has been used by several authors in poem generation task [10], [11], [12], [13]. Yan et al. [14] applied this model to couplet generation. They found that it can generate rather good couplets, and an additional polishing procedure can improve the quality even further.

Despite of the studies mentioned above, there are few couplet generation systems alive online, and no couplet generation systems are open-source, as far as we know. In this paper, we describe our VV-Couplet system. This system is also based on the sequence-to-sequence model with attention, as the work described by Yan et al. [14], but two innovations were introduced: (1) A local attention mechanism is introduced. Compare to the global attention used by Yan, it enforces the locality of the tonal and semantic constrains in couplets; (2) Entities, like person names and addresses, are carefully treated, which makes the generated couplets more compliant with rules. The VV-Couplet system has been public online[2], and we also released the source code as well as the training data[3].

II. Related work

Besides of the works mentioned above, we noticed some other research in couplet generation. Pan et al. [15] described a couplet generation system named as 'easyCouplet', which can generate not only couplet responses (the subsequent lines), but also couplet proposals (the antecedent lines). Lee et al. [16] built a web application that helps amateurs to produce Chinese poems. This system is mostly rule-based.

Some studies about poetry generation are also related to our work, in particular the neural-based methods. Zhang et al. [17] presented the first neural-based Chinese poetry generation system. Wang et al. [11] presented the first attention-based Chinese poetry generation system. Wang et al. [11] presented a planning network to enforce theme coherence, and Zhang et al. [13] proposed a memory-augmented architecture that can generate more creative poems.

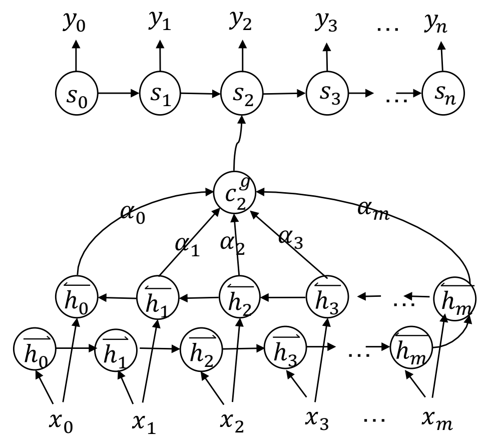

Fig. 4. Architecture of the sequence-to-sequence model with global attention.

III. Architecture

This section presents the neural architecture of VV-Couplet. We first review the vanilla attention-based sequence-to-sequence model, and then describe our local attention mechanism and entity treatment.

A. Attention-based Sequence to Sequence

The Sequence-to-Sequence model was first introduced by Sutskever et al. [8], where the input sequence is compressed into a fixed-length vector (encode) that is subsequently expanded to the output sequence (decode). This model has been used in numerous NLP tasks with great success, including machine translation and Chinese poem generation [7], [11]. This fixed-length encoding, however, makes it difficult to cope with long sentences. To address this issue, Bahdanau et al. [9] proposed an attention mechanism to generate the context dynamically. This attention-based sequence-to-sequence model has been regarded as state-of-art and successfully applied to many tasks, such machine translation, image caption, poem generation, etc.

The attention-based sequence to sequence model follows an encoder-decoder architecture, shown in Fig. 4. The encoder is a bi-directional recurrent neural network (RNN) with Long Short-Term Memory (LSTM) units that embeds the input a word sequence $[x_1, x_2, ..., x_m]$ to a sequence of hidden states $[h_1, h_2, ..., h_m]$, where each hidden state involves a forward state and a backward state, i.e., $h_i = [\overrightarrow{h_i}, \overleftarrow{h_i}]$. The decoder is another RNN that generates the target word sequence $[y_1, y_2, ..., y_n]$. To guide the generation to pay varied attention on source words at each step, an attention mechanism is introduced. Specifically, when generating the i-th target word, the attention paid on the j-th source word is measured by the relevance between the current hidden state of the decoder,

[1] http://duilian.msra.cn/
[2] http://139.199.22.149:82/
[3] https://gitlab.com/feng-7/VV-couplet.git

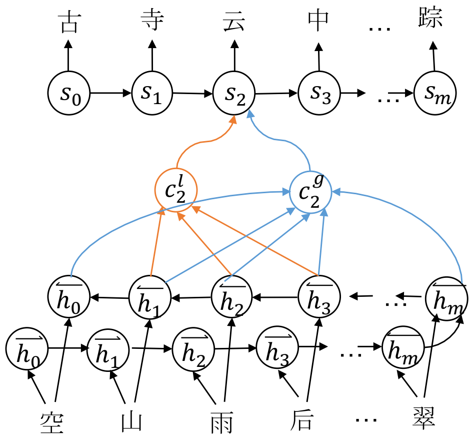

Fig. 5. Architecture of the sequence-to-sequence model with both global and local attention.

s_{i-1}, and the hidden state of the encoder at the j-th word, h_j, given by:

$$e_{ij} = a(s_{i-1}h_j); \qquad \alpha_{ij} = \frac{e_{ij}}{\sum_k e_{ik}}, \tag{1}$$

where $a(.,.)$ is the MLP-based relevance function, α_{ij} is the attention paid on x_j at the i-th decoding step. Since the attention will be paid on every source word, we call it *global attention mechanism*. By this global attention, the semantic context is global as well, which means that the context contains information from the whole source sentence. The context can be defined as a context vector, given as follows:

$$c_i^g = \sum_{j=1}^{m} \alpha_{ij} h_j. \tag{2}$$

Based on the context vector c_i^g, the state of the decoder in the previous step s_{i-1}, and the prediction of the previous step y_{i-1}, the decoder can generate a new word y_i and update the decoder state as follows:

$$p(y_i) = \sigma(y_i^T W s_i) \tag{3}$$

$$s_i = f_d(s_{i-1}, y_{i-1}, c_i^g) \tag{4}$$

where f_d is the recurrent function of the decoder, W is a projection matrix.

B. Local Attention Mechanism

The global attention model described above is a good Chinese poetry generation model, and can be applied to couplet generation directly, as reported by [14]. However, this model can be improved by considering the specialities of couplets. More specifically, the Chinese characters (or words) at *the same position* of the two lines should be closely related: inverse in tone and identical in lexical category. This means that the 'coupling rule' is rather local. This locality of the coupling rule has been summarized into some interesting jargons, such as '天对地 (Sky to Earth)', '雨对风 (Rain to Wind)', ' 仙鹤对神龙 (Crane to Dragon)', etc.

Inspired by the local adherence property, we hypothesize that when generating a subsequent line, more attention should be paid on characters at or near the position of the decoder in the antecedent line. However, the *global attention mechanism* used in vanilla attention-based model cannot take this feature into consideration. Therefore, we introduce a *local attention mechanism* to emphasize this local adherence between the antecedent and subsequent lines.

Fig. 5 shows the local-attention architecture. Besides of the global context vector c^g, the decoder also takes a local context vector c^l as an input. The update of decoder's hidden state changes to:

$$s_i = f_d(s_{i-1}, y_{i-1}, c_i^g, c_i^l). \tag{5}$$

Different from the global attention that is distributed to all the source words, the local attention is paid to words in local a window. Assuming the window is in the width of $2d + 1$, the local context vector at decoding step i is calculated by:

$$c_i^l = \sum_{j=i-d}^{i+d} \alpha_{ij} h_j, \tag{6}$$

where α_{ij} is computed in the same way as Equation (1). Note that the parameters used in local attention could be independent or shared with global attention. In this study, we choose shared parameters.

C. Entity Treatment

When generating couplets, an entity in the first line should be coupled with another entity in the second line at the same position. However, the neural model reads and generates character by character, which makes it difficult to recognize entities, i.e., person names or addresses. Additionally, most entities are low-frequent patterns and seldom appear in the training data. This means that entities cannot be well modelled by the sequence-to-sequence model, and their appearance will inevitably deteriorate the quality of the generated couplet.

We introduce a simple but effective strategy to address the entity issue. First, we collect a large entity database where the entities are grouped in categories. Second, entities in the antecedent line are recognized and removed, and the positions of the removed entities are recorded. Third, subsequent line of the entity-free antecedent line is generated as usually. Forth, the coupling entity is selected from the entity database for each removed entity, considering its category. Fifth, the selected coupling entity is inserted into the generated subsequent line at the position of the removed one. We find this simple approach can address most entities pretty well.

IV. Experiments

In this section, we will first present several baselines and implementation details of our local attention model, then describe the dataset used in experiments. The human evaluation process and the results will be presented subsequently.

A. Systems

Micro: This is the public service from Microsoft Research Asia[4]. We use the online results as our first baseline.

[4] http://duilian.msra.cn/

$Moses$: Moses is a state-of-art statistical machine translation (SMT) toolkit [18]. It is known to be good at learning word and phrase mappings, which is especially useful when generating couplets. Hence we take it as the second baseline. Note that the $Micro$ baseline is also based on SMT, but the details of the implementation is not known.

$Seq2Seq^g$: This is the attention-based sequence-to-sequence model with global attention. The implementation is based on the work by Wang et al. [11] and Zhang et al. [13]. Except the implementation details, the main architecture is the same as Yan's work [14] .

$Seq2Seq^{g+l}$: This is the model proposed in this paper, shown in Fig. 4, that combines global and local attention and applies the entity treatment strategy.

Implement details of the neural models ($Seq2Seq^g$ and $Seq2Seq^{g+l}$): the vocabulary (Chinese character) size is 6,493; the number of hidden units is 500; the window width of location attention is 3. The batch size is 80 during training, and greedy search is used during decoding.

B. Dataset

Our database is built using two sources. First, we took advantage of a public dataset[5] that contains about 700k couplets; Second, we collected about 80k couplets from the Internet. These two datasets are merged into a large database that contains 780k couplets. We chose 784,975 couplets as the training set and 400 couplets as the test set.

C. Evaluation

We invited 12 experts to participate in the evaluation process, and all of them have rich experience not only in evaluating couplets, but also in writing them. Most of the experts are from prestigious institutes, including Peking university and the Chinese Academy of Social Science (CASS). Following the suggestions of the experts, we use two metrics to evaluate the generation, as listed below:

- $Semantic\ Coherrence$: if a couplet reads fluently, conveys reasonable meaning, and has any aesthetic feeling.
- $Rule\ Compliance$: if a couplet strictly follows the coupling rule, i.e., 'inverse in tone and identical in lexical category'.

We designed an online survey. Each questionnaire contained 15 questions. Each question presented one antecedent line in the testing set and four subsequent lines generated from the four systems respectively. Evaluators were asked to choose the best one from the four choices in terms of the two metrics mentioned above. Finally, we got 30 completed questionnaires, i.e. 450 questions. Every example in the test set has been evaluated at least once.

D. Results

The results of the human evaluation are shown in Table I, where the number means in how many questions the model's generation was selected as the best. It can be observed that our model, $Seq2Seq^{g+l}$ performs the best in terms of both $Semantic\ Coherrence$ and $Rule\ Compliance$. Especially, the better performance of $Seq2Seq^{g+l}$ compared to $Seq2Seq^g$ confirms the effectiveness of local attention, indicating that emphasis on local adherence is important for couplet generation. Additionally, the comparison between $Seq2Seq^g$ and $Moses$ shows that neural models tend to learn coherent and aesthetic sequence, while SMT models are more good at learning coupling rules. A key contribution of the local attention mechanism is that it amends the shortcoming of neural models in coupling rules, hence generating couplets fluent, coherent, aesthetic and rule-complied. Note that all the systems we built outperform $Micro$ in a significant way. This should be attributed to the large training data used in our study.

[5]https://github.com/wb14123/couplet-dataset

TABLE I
COUPLET GENERATION PERFORMANCE

System	Semantic Coherrence	Rule Compliance
$Micro$	54	55
$Moses$	111	128
$Seq2Seq^g$	134	126
$Seq2Seq^{g+l}$	**151**	**141**

	雪后放晴清一色 The world is pure white when it is clearing up after the snow.
$Micro$	风前落雨玉无边 The world is boundlessly jade when it is raining before the wind.
$Moses$	花前说爱结同心 We promise to love each other forever in front of flowers.
$Seq2Seq^g$	春初开泰喜千家 Thousands of families are enjoying the happiness of early Spring.
$Seq2Seq^{g+l}$	风前吹雨绿三分 The world is light green when it is drizzling before the wind.

Fig. 6. An example of couplets that generated by four systems.

Fig. 6 gives an example to illustrate the characteristics of the four systems. Given the same antecedent line '雪后放晴清一色', $Seq2Seq^{g+l}$ generated the best subsequent line. Its success comes from two aspects: First, it demonstrates better local coherence: almost every word is perfectly coupled its counterpart in the antecedent line, e.g., '风前 (before the wind)' to '雪后 (after the snow)', '吹雨 (drizzling)' to '放晴 (clearing up)', '清一色 (pure white)' to '绿三分 (light green)'; Second, it shows good semantic fluency: the antecedent line describes the white scenery after snowing, and the subsequent line describes a similar scenery. In comparison, the result from $Moses$ talks a different theme (human love), leading to reduced coherence. The result from $Seq2Seq^g$ is fine in theme, but the coupling is not as good, e.g., '春(spring)' is not well coupled with '雪(snow)', '清(clean)' is not well coupled with '喜(happy)'. The result from $Micro$ looks fine in character coupling, but the phrase '玉无边 (boundlessly jade)' does not make sense in Chinese.

Fig. 7 gives an example to demonstrate the entity treatment. We use the example shown in Fig. 3, and generate the subsequent line using the $Seq2Seq^{g+l}$ system with and without(w/o) the entity treatment. It can be seen that with

	董卿观花花开花落花容晚 Dong Qing perceives flowers that bloom and fall
Oracle	曹颖望月月隐月圆月色浓 Cao Ying gazes moon that appears and absents
With Entity Treatment	朱迅望月月圆月缺月色明 Zhu Xun gazes moon that appears and fades
W/O Entity Treatment	秋风送雨雨去雨来雨意浓 Autumn wind accompanies rain that drops and stops

Fig. 7. An example of entity treatment.

Fig. 8. The WeChat mini-program of VV-Couplet.

the entity treatment, '朱迅 (Zhu Xun)' is selected to couple with '董卿 (Dong Qing)', both are famous compere names; otherwise the system will generate '秋风 (autumn wind)', which is clearly a bad case.

E. Online Service

Finally, we released a WeChat mini-program, shown in Fig 8, and a online web service, shown in Fig 9. The source code plus the training data will be also published.

V. Conclusions

We described VV-Couplet, an open source couplet generation system. The core of VV-couplet is a sequence-to-sequence neural model, accompanied by a local attention mechanism. Our experimental results demonstrated that this architecture can learn both local coupling rules and global semantic meaning, and exhibited better subjective scores in both rule-compliance and semantic coherence when compared with SMT-based methods and the global-attention counterpart. An interesting work that we are working on is to test if this system can pass the Turing test, and the results seem promising.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Projects 61633013 and 61371136.

Fig. 9. The website of VV-Couplet.

References

[1] Y. Yi, Z. He, L. Li, J. Zhou, Y. Qu, and H. Zhang, "On computation models of Chinese couplet responses," *Computer Science*, vol. 33, no. 4, pp. 162–163, 2006.

[2] K.-x. ZHANG and M.-s. SUN, "An Chinese couplet generation model based on statistics and rules," *Journal of Chinese Information Processing*, vol. 1, p. 017, 2009.

[3] M. Zhou, L. Jiang, and J. He, "Generating Chinese couplets and quatrain using a statistical approach," in *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation, Volume 1*, vol. 1, 2009.

[4] L. Jiang and M. Zhou, "Generating Chinese couplets using a statistical mt approach," in *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2008, pp. 377–384.

[5] J. He, M. Zhou, and L. Jiang, "Generating Chinese classical poems with statistical machine translation models." in *AAAI*, 2012.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[7] Y. Feng, S. Zhang, A. Zhang, D. Wang, and A. Abel, "Memory-augmented neural machine translation," *arXiv preprint arXiv:1708.02005*, 2017.

[8] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[9] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *ICLR'2015*, 2015.

[10] Z. Wang, W. He, H. Wu, H. Wu, W. Li, H. Wang, and E. Chen, "Chinese poetry generation with planning based neural network," *arXiv preprint arXiv:1610.09889*, 2016.

[11] Q. Wang, T. Luo, D. Wang, and C. Xing, "Chinese song iambics generation with neural attention-based model," *arXiv preprint arXiv:1604.06274*, 2016.

[12] X. Yi, R. Li, and M. Sun, "Generating Chinese classical poems with rnn encoder-decoder," in *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*. Springer, 2017, pp. 211–223.

[13] J. Zhang, Y. Feng, D. Wang, Y. Wang, A. Abel, S. Zhang, and A. Zhang, "Flexible and creative Chinese poetry generation using neural memory," *arXiv preprint arXiv:1705.03773*, 2017.

[14] R. Yan, C.-T. Li, X. Hu, and M. Zhang, "Chinese couplet generation with neural network structures," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2016, pp. 2347–2357.

[15] Z. Pan, S. Zhang, and Y. Guo, "Easycouplet: Automatic generation of Chinese traditional couplets," in *Transactions on Edutainment XIV*. Springer, 2018, pp. 117–132.

[16] J. Lee, Y. C. Hui, and Y. H. Kong, "Knowledge-rich, computer-assisted composition of Chinese couplets," *Literary and Linguistic Computing*, vol. 31, no. 1, pp. 152–163, 2014.

[17] X. Zhang and M. Lapata, "Chinese poetry generation with recurrent neural networks," in *EMNLP 2014*, 2014, pp. 670–680.

[18] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens *et al.*, "Moses: Open source toolkit for statistical machine translation," in *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions*. Association for Computational Linguistics, 2007, pp. 177–180.