Automatic Poetry Generation with Mutual Reinforcement Learning

Xiaoyuan Yi[1], Maosong Sun[1], Ruoyu Li[2], Wenhao Li[1]

[1]Tsinghua University

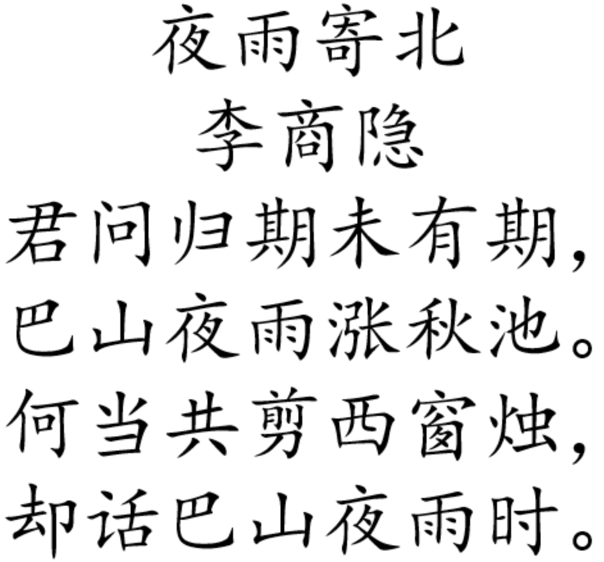
[2] 6ESTATES PTE LTD

EMNLP 2018

CONTENT

1. Background & Motivation
2. Single-Learner Reinforcement Learning
3. Mutual Reinforcement Learning
4. Experiments & Conclusion
5. Jiuge (九歌) System

Background & Motivation



夜雨寄北
李商隐
君问归期未有期，
巴山夜雨涨秋池。
何当共剪西窗烛，
却话巴山夜雨时。

There was a Young Person of Ayr,
Whose head was remarkably square:
On the top, in fine weather,
She wore a Gold Feather,
Which dazzled the people of Ayr.
——By Edward Lear

(1) Concise language

(2) Exquisite expression

(3) Rich content

Background & Motivation

Entertainments

Poetry Education

Literary Research

…

A desirable entry point for automatic analyzing, understanding and utilizing literary text.

Background & Motivation

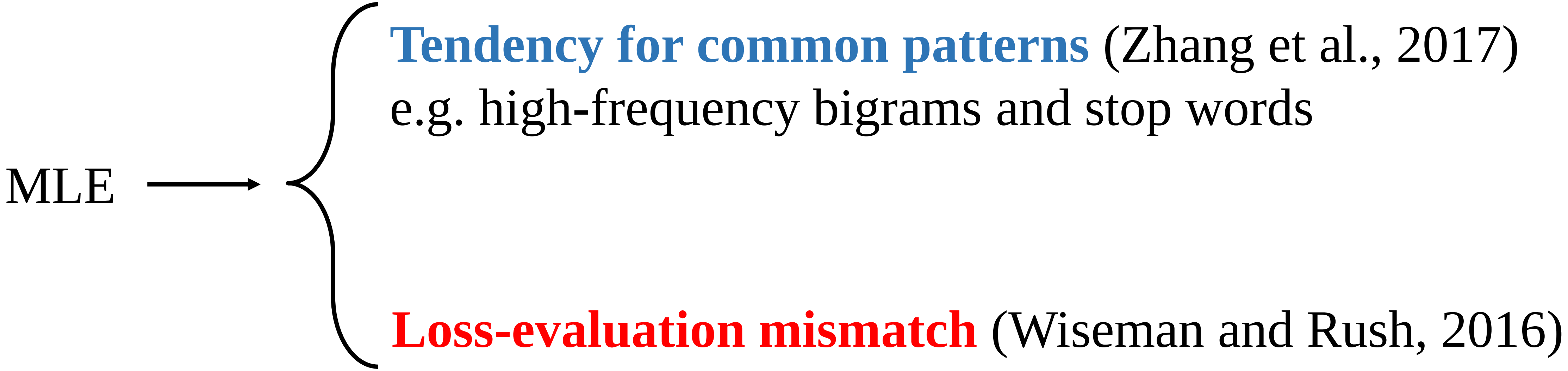
- Fluency (Zhang and Lapata, 2014)
- Coherence (Wang et al., 2016)
- Overall Quality (Yan, 2016)
- Meaningfulness (Ghazvininejad et al., 2016)
- Innovation (Zhang et al., 2017)

…

Maximum Likelihood Estimation (MLE)

?

Background & Motivation



MLE

Tendency for common patterns (Zhang et al., 2017)
e.g. high-frequency bigrams and stop words

Loss-evaluation mismatch (Wiseman and Rush, 2016)

Background & Motivation

MLE → **Loss-evaluation mismatch**

	MLE	Human
evaluation granularity mismatch	word-level loss	sequence level (a poem line) discourse level (a whole poem)
criterion mismatch	likelihood	some human criteria

Fluency Coherence

Meaningfulness Overall quality

Background & Motivation

- Further design more sophisticated model structures. 😔

- Directly model the human evaluation criteria and use them as explicit rewards to guide gradient update by reinforcement learning. 😄

CONTENT

1. Background & Motivation
2. Single-Learner Reinforcement Learning
3. Mutual Reinforcement Learning
4. Experiments & Conclusion
5. Jiuge (九歌) System

Single-Learner Reinforcement Learning

Input: K user keywords $W = \{w_k\}_{k=1}^{K}$

Output: A poem consisting of n lines $O = \{L_i\}_{i=1}^{n}$

A basic poetry generator: $P_g(\cdot\,|W;\theta)$

(pre-trained with MLE loss)

- **Fluency**: are the lines fluent and well-formed?
- **Coherence**: is the poem as a whole coherent in meaning and theme?
- **Meaningfulness**: does the poem convey some certain messages?
- **Overall quality**: the reader's general impression on the poem

Single-Learner Reinforcement Learning

Fluency Rewarder $R_1(O)$

$$r(L_i) = max(|P_{lm}(L_i - \mu)| - \delta_1 * \sigma, 0)$$

$$R_1(O) = \frac{1}{n}\sum_{i=1}^{n} e^{-r(L_i)}$$

Motivate the language model probability of generated lines to fall into a reasonable range.

Coherence Rewarder $R_2(O)$

$$MI(L_{1:i-1}, L_i) = logP_{seq2seq}(L_i|L_{1:i-1}) - \lambda logP_{lm}(L_i)$$

$$R_2(O) = \frac{1}{n-1}\sum_{i=2}^{n} MI(L_{1:i-1}, L_i)$$

Use Mutual Information to measure the coherence and expect higher MI.

Single-Learner Reinforcement Learning

Meaningfulness Rewarder $R_3(O)$

MLE-based models → common and meaningless words.

e.g., 不知 (bu zhi, don’t know) 何处 (he chu, where) 无人 (wu ren, no one)

(Similar issues arise in dialog generation task.)

$$R_3(O)=\frac{1}{n}\sum_{i=1}^{n}\boxed{F(L_i)}$$

- TF-IDF → more meaningful words
- A rough attempt but improves user experience.

- A neural network to estimate the TF-IDF value of a line.
- To tackle the OOV problem when sampling.

Single-Learner Reinforcement Learning

Overall Quality Rewarder $R_4(O)$

When evaluating, human experts:

- focus on discourse-level;
- ignore some minor defects;
- judge the overall quality of a whole poem.

$$R_4(O) = \sum_{k=1}^{3} \boxed{P_{cl}(k|O)} * k$$

A strong classifier to classify a poem into:
Class 1: computer-generated poetry
Class 2: ordinary human-authored poetry
Class 3: masterpieces

Motivates the generated poems to resemble masterpieces in a higher-level!

Single-Learner Reinforcement Learning

Final Reward:

$$\begin{tikzcd}
R(O) = \sum_{j=1}^{4} \alpha_j * R_j(O) \arrow[r, "\text{reduce variance}"] & R'(O) \arrow[d] \\
\text{Use REINFORCE to minimize:} & L_{DRL}(\theta) = -\sum_{m=1}^{M} E_{O\sim P_g(\cdot|W^m;\theta)}(R'(O)) \arrow[d, "\text{Combine MEL loss and DRL loss}"] \\
& L(\theta) = L_{MLE}(\theta) + \beta * L_{DRL}(\theta)
\end{tikzcd}$$

CONTENT

1. Background & Motivation
2. Single-Learner Reinforcement Learning
3. Mutual Reinforcement Learning
4. Experiments & Conclusion
5. Jiuge (九歌) System

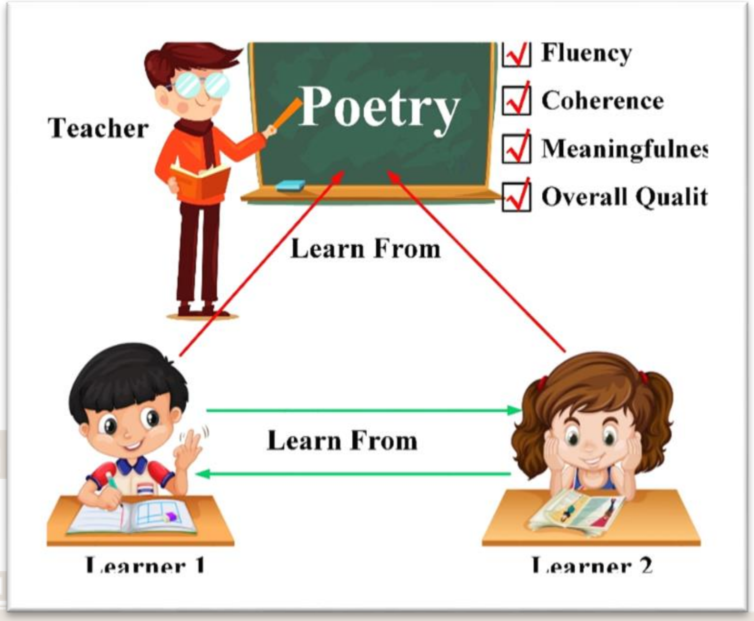
Mutual Reinforcement Learning

Previous models always treat literary text generation as a **solo (single-handed)** task. 🤔

In writing theories:
it is shown that writing is supported as an activity in which writers will learn from more experienced writers, such as other students, teachers, or authors (Prior, 2006).

Allow communication among learners (generators) 😃

During the training process, use two different generators and enable them to learn not only from the teacher (rewarder) but also from the other.



Mutual Reinforcement Learning

- Local Mutual Reinforcement Learning (S-MRL)

Define two generators as: $P_g(\theta_1)$ $P_g(\theta_2)$

```
For the same input keywords $W$:
$1. O_1 \sim P_g(\theta_1), \; O_2 \sim P_g(\theta_2);$
2. If R$(O_1) > R(O_2) * (1+\delta_2)$ and $R_j(O_1) > R_j(O_2)$ for all j:
        Update $\theta_1, \theta_2$ with $O_1$;
    else if R$(O_1) < R(O_2) * (1+\delta_2)$ and $R_j(O_1) < R_j(O_2)$ for all j:
        Update$\theta_1, \theta_2$ with $O_2$;
    else:
        Update $\theta_1$ with $O_1$; $\theta_2$ with $O_2$.
```

If a learner creates a significantly better poem,
then the other learner will learn it!

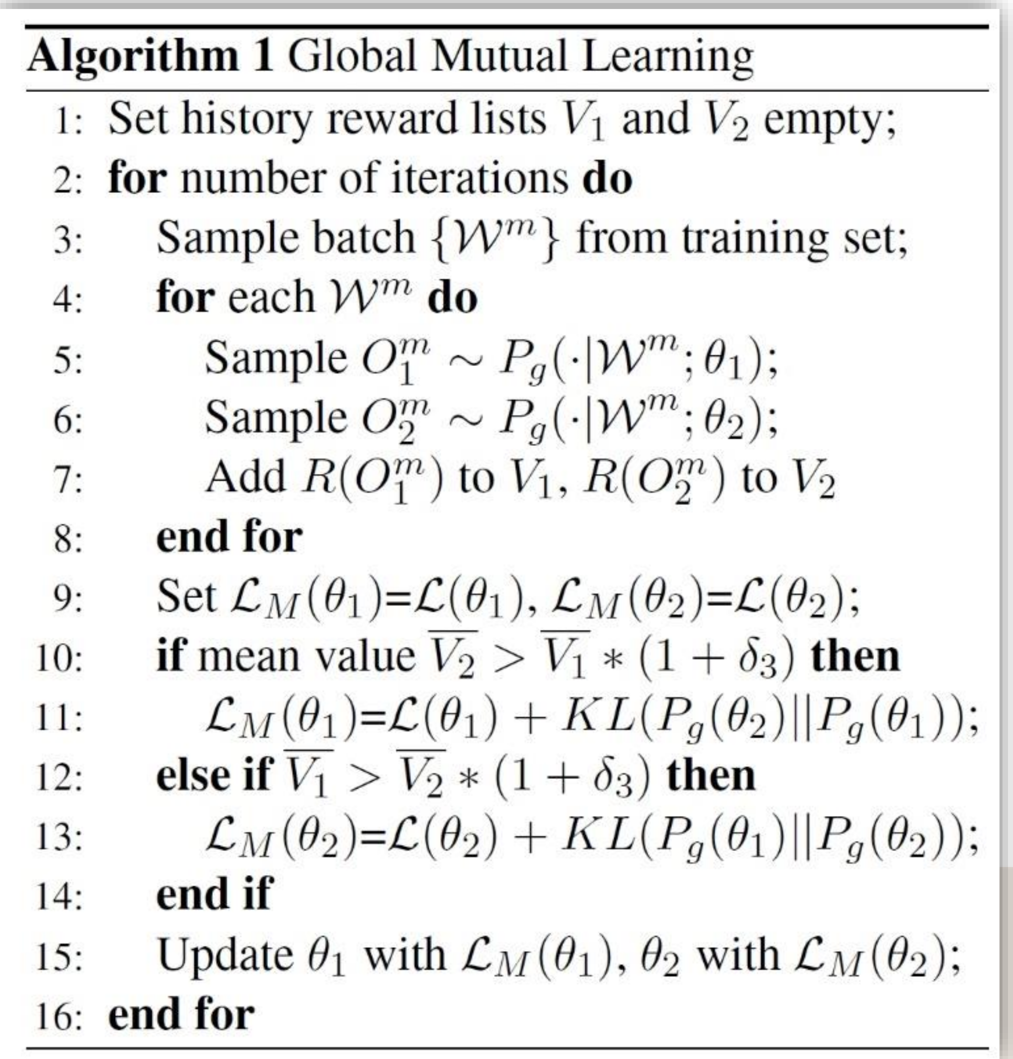
Mutual Reinforcement Learning

- Local Mutual Reinforcement Learning (S-MRL)

- is an instance-based method;
- gives a generator more high-reward instances;
- can be considered as searching the policy space along two different paths;
- allows the generators to explore larger space and find a more proper direction so as to escape from the local minima.

Mutual Reinforcement Learning

- Global Mutual Reinforcement Learning (G-MRL)



Algorithm 1 Global Mutual Learning

```
Set history reward lists $V_1$ and $V_2$ empty;
for number of iterations do
    Sample batch $\{\mathcal{W}^m\}$ from training set;
    for each $\mathcal{W}^m$ do
        Sample $O_1^m \sim P_g(\cdot|\mathcal{W}^m;\theta_1)$;
        Sample $O_2^m \sim P_g(\cdot|\mathcal{W}^m;\theta_2)$;
        Add $R(O_1^m)$ to $V_1$, $R(O_2^m)$ to $V_2$
    end for
    Set $\mathcal{L}_M(\theta_1)$=$\mathcal{L}(\theta_1)$, $\mathcal{L}_M(\theta_2)$=$\mathcal{L}(\theta_2)$;
    if mean value $\overline{V_2} > \overline{V_1} * (1+\delta_3)$ then
        $\mathcal{L}_M(\theta_1)$=$\mathcal{L}(\theta_1) + KL(P_g(\theta_2)||P_g(\theta_1))$;
    else if $\overline{V_1} > \overline{V_2} * (1+\delta_3)$ then
        $\mathcal{L}_M(\theta_2)$=$\mathcal{L}(\theta_2) + KL(P_g(\theta_1)||P_g(\theta_2))$;
    end if
    Update $\theta_1$ with $\mathcal{L}_M(\theta_1)$, $\theta_2$ with $\mathcal{L}_M(\theta_2)$;
end for
```

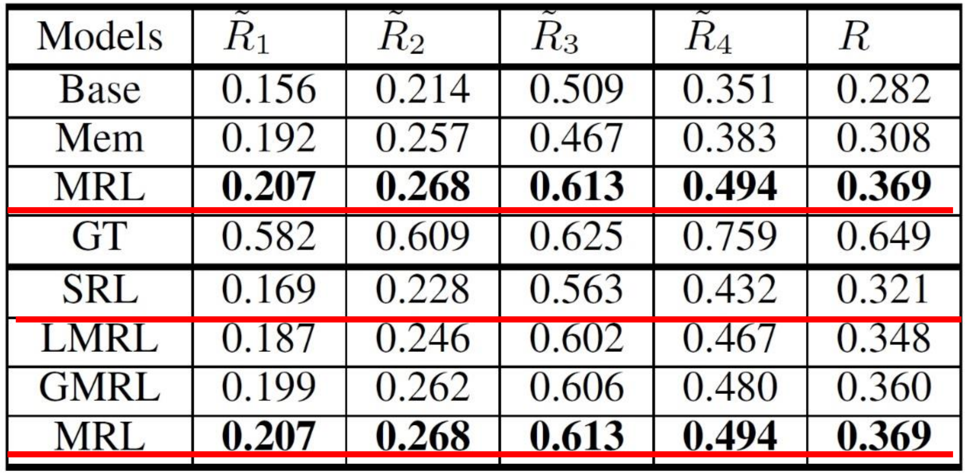
- is an distribution-level method;
- takes long-period history into account;
- pulls the distribution towards the better one of the two generators.

Local MRL + Global MRL

CONTENT

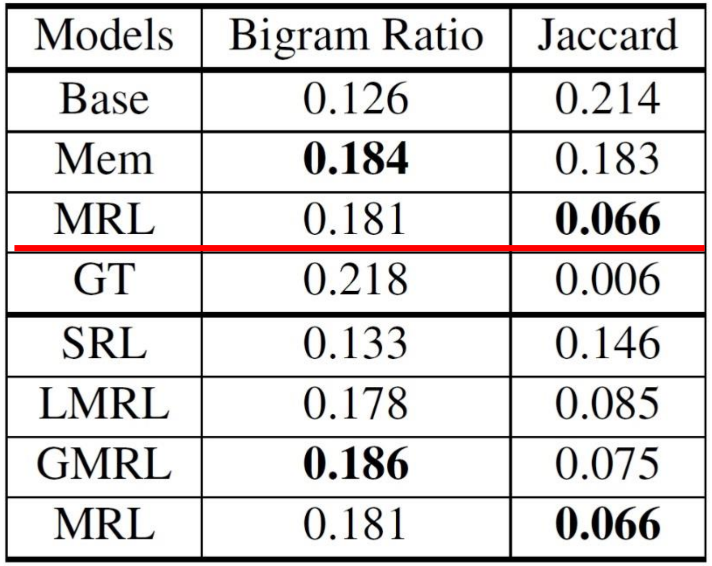
1. Background & Motivation
2. Single-Learner Reinforcement Learning
3. Mutual Reinforcement Learning
4. Experiments & Conclusion
5. Jiuge (九歌) System

Experiments & Conclusion



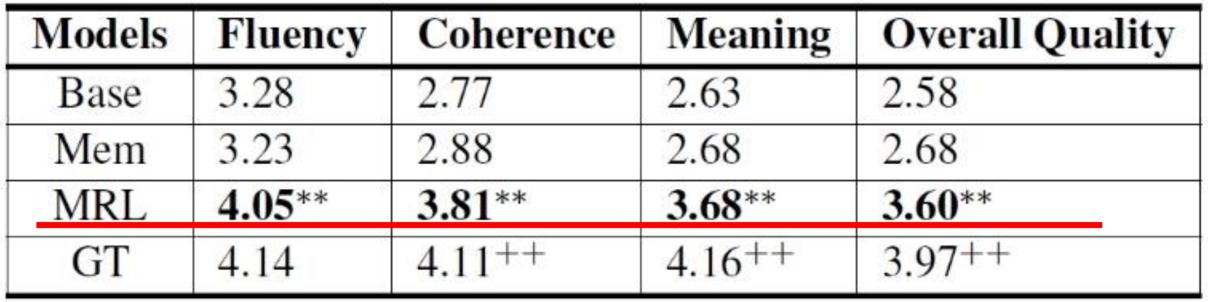
Models	$\tilde{R}_1$	$\tilde{R}_2$	$\tilde{R}_3$	$\tilde{R}_4$	R
Base	0.156	0.214	0.509	0.351	0.282
Mem	0.192	0.257	0.467	0.383	0.308
MRL	**0.207**	**0.268**	**0.613**	**0.494**	**0.369**
GT	0.582	0.609	0.625	0.759	0.649
SRL	0.169	0.228	0.563	0.432	0.321
LMRL	0.187	0.246	0.602	0.467	0.348
GMRL	0.199	0.262	0.606	0.480	0.360
MRL	**0.207**	**0.268**	**0.613**	**0.494**	**0.369**

Table 1: Automatic rewards of different models and strategies. $\tilde{R}_1$: fluency, $\tilde{R}_2$: coherence, $\tilde{R}_3$: meaningfulness, $\tilde{R}_4$: overall quality, R: weighted-average reward. LMRL: local MRL, GMRL: global MRL.



Models	Bigram Ratio	Jaccard
Base	0.126	0.214
Mem	**0.184**	0.183
MRL	0.181	**0.066**
GT	0.218	0.006
SRL	0.133	0.146
LMRL	0.178	0.085
GMRL	**0.186**	0.075
MRL	0.181	**0.066**

Table 2: Automatic evaluation results of diversity and innovation. The Jaccard values are multiplied by 10 for clearer observation. We expect higher bigram ratio and smaller Jaccard values.



Models	**Fluency**	**Coherence**	**Meaning**	**Overall Quality**
Base	3.28	2.77	2.63	2.58
Mem	3.23	2.88	2.68	2.68
MRL	**4.05****	**3.81****	**3.68****	**3.60****
GT	4.14	4.11^{++}	4.16^{++}	3.97^{++}

Table 3: Human evaluation results. Diacritic ** ($p < 0.01$) indicates MRL significantly outperforms baselines; ++ ($p < 0.01$) indicates GT is significantly better than all models.

Experiments & Conclusion

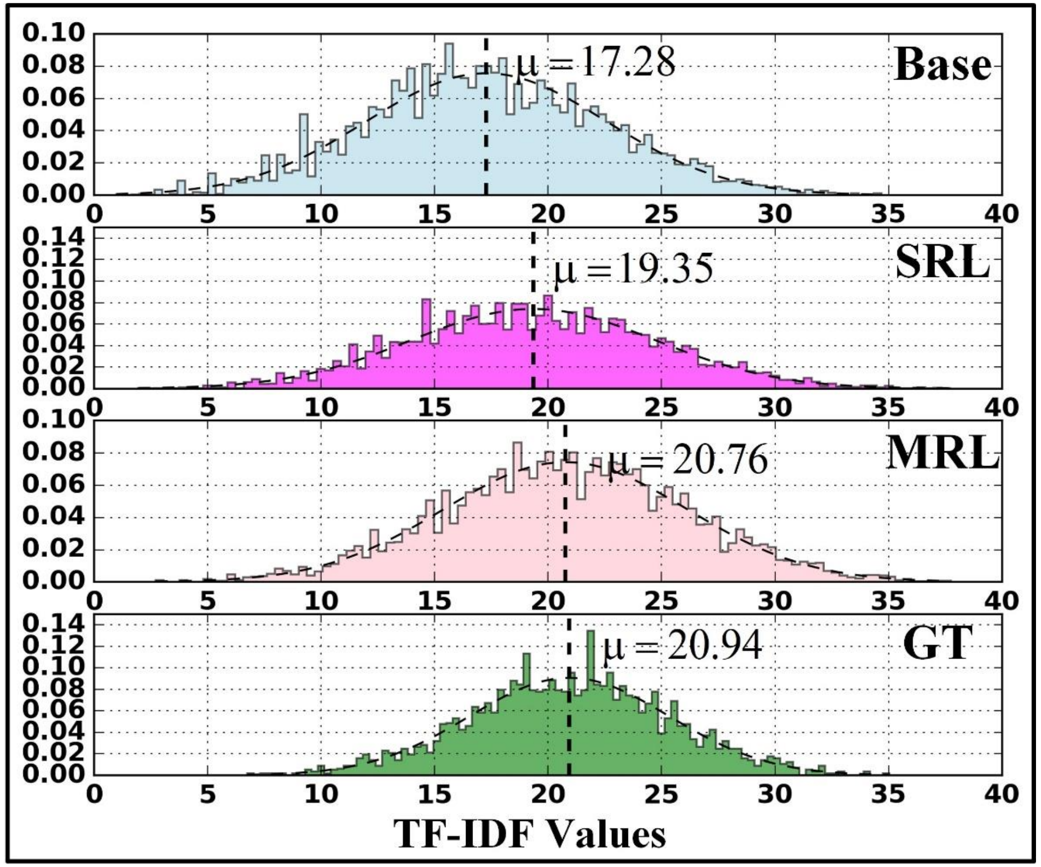


Figure 1: TF-IDF distributions of poems generated by different models. We show real TF-IDF, instead of the estimated value.

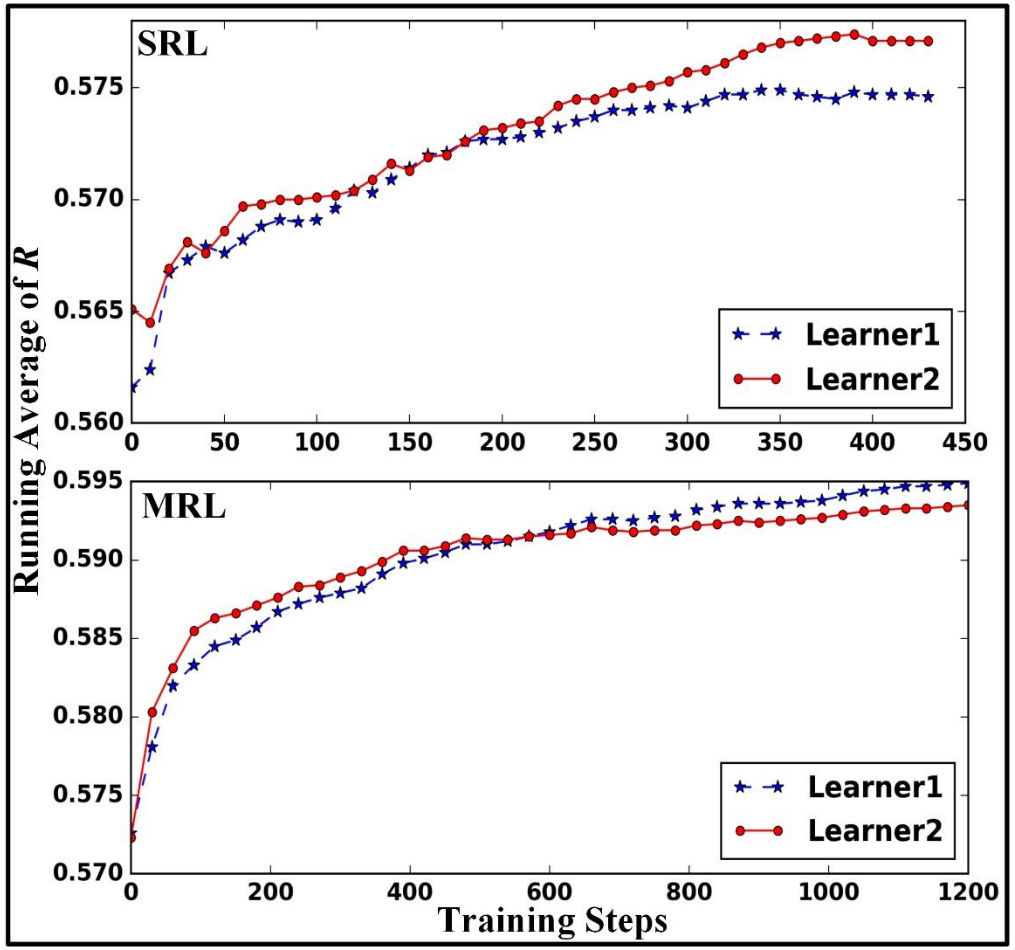
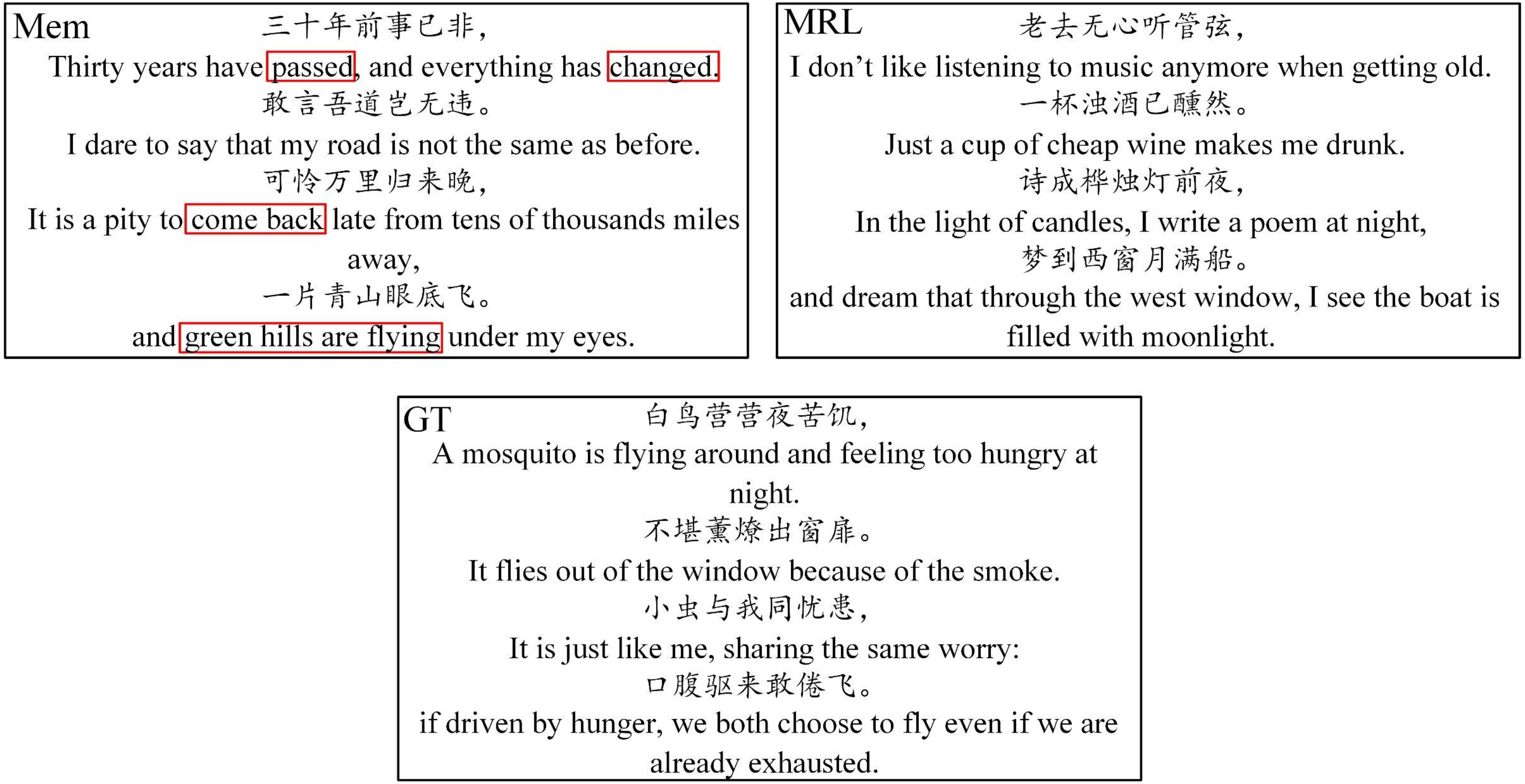


Figure 2: The learning curves of SRL and MRL. Learner 2 (red-dotted line) is a better pre-trained generator. Learner 1 (blue-star line) is a not so good pre-trained generator

Experiments & Conclusion



Mem

三十年前事已非，
Thirty years have passed, and everything has changed.
敢言吾道岂无违。
I dare to say that my road is not the same as before.
可怜万里归来晚，
It is a pity to come back late from tens of thousands miles away,
一片青山眼底飞。
and green hills are flying under my eyes.

MRL

老去无心听管弦，
I don't like listening to music anymore when getting old.
一杯浊酒已醺然。
Just a cup of cheap wine makes me drunk.
诗成桦烛灯前夜，
In the light of candles, I write a poem at night,
梦到西窗月满船。
and dream that through the west window, I see the boat is filled with moonlight.

GT

白鸟营营夜苦饥，
A mosquito is flying around and feeling too hungry at night.
不堪薰燎出窗扉。
It flies out of the window because of the smoke.
小虫与我同忧患，
It is just like me, sharing the same worry:
口腹驱来敢倦飞。
if driven by hunger, we both choose to fly even if we are already exhausted.

Figure 3: Sampled poems generated (with the same input keywords) by different models: Mem (Zhang et al., 2017), MRL (our model), GT (ground-truth, human –authored poem). Some defects are shown in red boxes.

Experiments & Conclusion

- First utilize reinforcement learning to generate poetry
  - Directly model and optimize human evaluation criteria.
  - Alleviate the loss-evaluation mismatch problem in poetry generation.
- Mutual Reinforcement Learning
  - Writing theory motivation
  - A step towards multi-agent DRL in literary text generation.
  - Treat automatic writing as a communication-involved process to further improve performance.
- Prominent improvement on Chinese poetry, outperforming the state-of-the-art model.

CONTENT

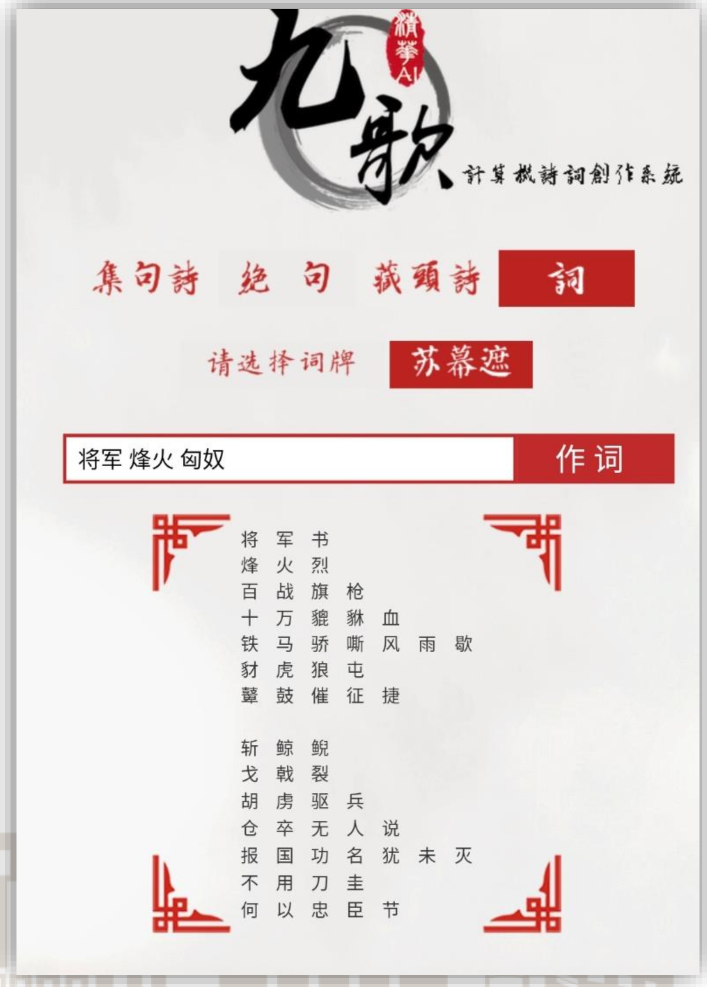
1. Background & Motivation
2. Single-Learner Reinforcement Learning
3. Mutual Reinforcement Learning
4. Experiments & Conclusion
5. Jiuge (九歌) System

Jiuge (九歌), a Chinese poetry generation system developed by THU NLP&CSS lab.

- Support most popular genres of Chinese poetry
- Online generation interface
- Page View > 2 million

The proposed model will be integrated into Jiuge!

https://jiuge.thunlp.cn/



将军 烽火 匈奴

将军书
烽火烈
百战旗枪
十万貔貅血
铁马骄嘶风雨歇
豺虎狼屯
鼙鼓催征捷

斩鲸鲵
戈戟裂
胡虏驱兵
仓卒无人说
报国功名犹未灭
不用刀圭
何以忠臣节

Poster Presentation by Cheng Yang (09:00, 4 Nov, Grand Hall 2):

Stylistic Chinese Poetry Generation via Unsupervised Style Disentanglement

https://jiuge.thunlp.cn/

Thanks!

Xiaoyuan Yi

THUNLP&CSS Lab, Tsinghua University

Mail: yi-xy@mails.tsinghua.edu.cn

https://jiuge.thunlp.cn/

Thanks!

Xiaoyuan Yi

THUNLP&CSS Lab, Tsinghua University

Mail: yi-xy@mails.tsinghua.edu.cn

九歌 清華AI

https://jiuge.thunlp.cn/