Research Progress on Story Generation

Wei Wang[1,2]

1. Graduate School at Shenzhen, Tsinghua University

2. Tencent AI Lab, NLP Center

2019.07

Overview

- A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation(EMNLP 2018)
- Plan-And-Write: Towards Better Automatic Storytelling(AAAI 2019)
- Strategies for Structuring Story Generation(ACL 2019)

Problem Definition

- Input: a prompt
- Output: a story
- Key aspects: Fluency, relevance, coherence

- These three paper all use Intermediate representation:
  - Key phrases
  - Keywords
  - SRL

A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation

Jingjing Xu[1], Xuancheng Ren[1], Yi Zhang[1], Qi Zeng[1], Xiaoyan Cai[2], Xu Sun[1]
[1]MOE Key Lab of Computational Linguistics, School of EECS, Peking University
[2]School of Automation, Northwestern Polytechnical University
{jingjingxu,renxc,zhangyi16,pkuzengqi,xusun}@pku.edu.cn
xiaoyanc@nwpu.edu.cn

Introduction

A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation(EMNLP 2018)

- The connection among sentences is mainly reflected through key phrases, such as predicates, subjects, objects and so on. The other words (e.g., modifiers) not only are redundant for understanding semantic dependency, but also make the dependency sparse.

Task Description
Input: A short description of a scene or an event. Output: A relevant narrative story following the input.
Examples
Input: *Fans came together to celebrate the opening of a new studio for an artist.* **Output**: *The artist provided champagne in flutes for everyone. Friends toasted and cheered the artist as she opened her new studio.*
Input: *Last week I attended a wedding for the first time.* **Output**: *There were a lot of families there. They were all taking pictures together. Everyone was very happy. The bride and groom got to ride in a limo that they rented.*

Table 1: An illustration of narrative story generation.

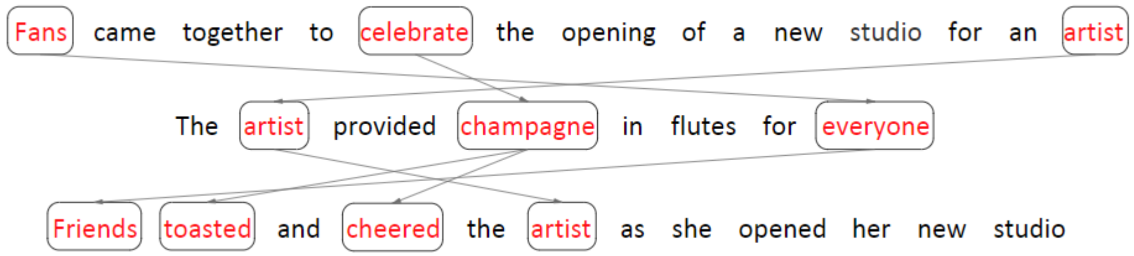
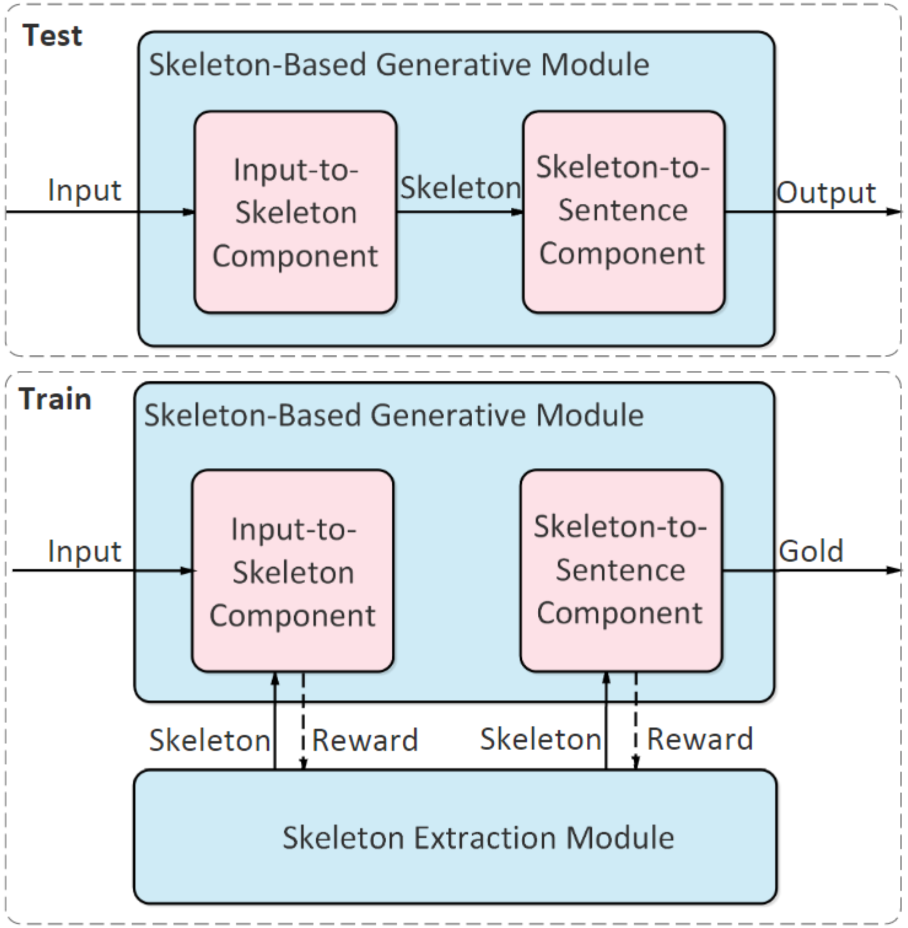


Figure 1: The semantic dependency among sentences in a narrative story. It can be seen that the connection among sentences is mainly reflected through key phrases (shown in red). In this work, we regard such key phrases as a skeleton.

Method

A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation(EMNLP 2018)

- A skeleton-based generative module.
  - An input-to-skeleton component.
    - A Seq2Seq structure with a hierarchical encoder and an attention-based decoder.
  - A skeleton-to-sentence component.
    - A Seq2Seq structure. Both the encoder and the decoder are one-layer LSTM networks with the attention mechanism.
- A skeleton extraction module.
    - A Seq2Seq structure.



Method

A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation(EMNLP 2018)

- A skeleton-based generative module.
  - An input-to-skeleton component.

$$L_\alpha = -\sum_{i=1}^{T} P_Q(s_i|\boldsymbol{c}, \alpha) \tag{1}$$

  - A skeleton-to-sentence component.

$$L_\theta = -\sum_{i=1}^{M} P_D(y_i|\boldsymbol{s}, \theta) \tag{2}$$

- A skeleton extraction module.

$$L_\gamma = -\sum_{i=1}^{T} P_E(s_i|\boldsymbol{x}, \gamma) \tag{3}$$

Algorithm 1 The reinforcement learning method for training the generative module G_ϕ and the skeleton extraction module E_γ.

1: Initialize the generative module G_ϕ and the skeleton extraction module E_γ with random weights ϕ, γ
2: Pre-train E_γ using MLE based on Eq. 3
3: **for** each iteration $j = 1, 2, ..., J$ **do**
4: Generate a skeleton $\boldsymbol{s}_j$ based on E_γ
5: Given $\boldsymbol{s}_j$, train G_ϕ based on Eq. 1 and Eq. 2.
6: Compute the reward R_c based on Eq. 5
7: Compute the gradient of E_γ based on Eq. 4
8: Update the model parameter γ
9: **end for**

$$R_c = [K - (R_1 \times R_2)^{\frac{1}{2}}] \tag{5}$$

$$\nabla J(\gamma) = \mathbb{E}[R_c \cdot \nabla \log(P_E(\boldsymbol{s}|\boldsymbol{x}), \gamma)] \tag{4}$$

Experiment

A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation(EMNLP 2018)

- The differences between the existing state-of-the-art models are within 0.07, while the proposed model supersedes the best of them by 0.13.
- In terms of coherence, the proposed model is better than all the existing models.
- Overall, the proposed model is arguably better than the existing models in that it achieves a balance between coherence and fluency.
- The generalized templates can constrain the search space in generation and the model achieves higher fluency by loss of expressive power. GE-Seq2Seq model does not learn the dependency among sentences effectively.

Models	BLEU
EE-Seq2Seq	0.0029
DE-Seq2Seq	0.0027
GE-Seq2Seq	0.0022
Proposed Model	**0.0042 (+44.8%)**

Table 2: Automatic evaluations of the proposed model and the state-of-the-art models.

Models	Fluency	Coherence	G-Score
EE-Seq2Seq	6.28	5.14	5.68
DE-Seq2Seq	8.48	3.54	5.48
GE-Seq2Seq	9.48	3.58	5.82
Proposed Model	8.69	5.62	**6.99 (+20.1%)**

Table 3: Human evaluations of the proposed model and the state-of-the-art models.

Experiment

A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation(EMNLP 2018)

- With reinforcement learning, the BLEU score significantly overpasses the Seq2Seq model by 40%.
- The slight improvement with the skeleton extraction module in BLEU reflects as the decreases in both fluency and coherence.
- The style of the dataset for pre-training the skeleton extraction module is very different from the narrative story dataset.
- The Seq2Seq model beats the existing state-of-the-art models (DE-Seq2Seq and GE-Seq2Seq) in human evaluation and automatic evaluation. It is mainly attributed to the oversimplification of sentences.

Models	BLEU
Seq2Seq	0.0028
+Skeleton Extraction Module	0.0029
+Reinforcement Learning	**0.0042**

Table 5: Automatic evaluations of key components.

Models	Fluency	Coherence	G-Score
Seq2Seq	7.54	4.98	6.13
+Skeleton Extraction Module	7.26	4.32	5.60
+Reinforcement Learning	8.69	5.62	**6.99**

Table 6: Human evaluations of the key components.

Experiment

A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation(EMNLP 2018)

- It is expected that the irrelevant scenes make up most of the errors.
- In addition, there are several examples that are hard to be understood due to chaotic syntax.
- For the type of chaotic timeline, the model neglects the time order of scenes and the generated stories goes backward in time.
- The repeated scenes mean that the generated stories just describe the input again.

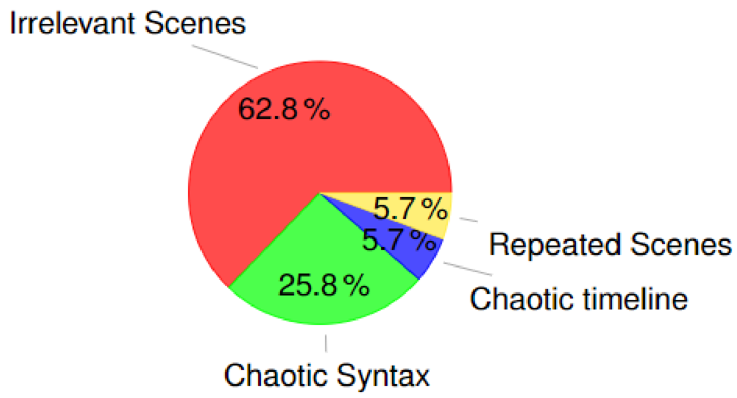


Figure 3: The distribution of error types.

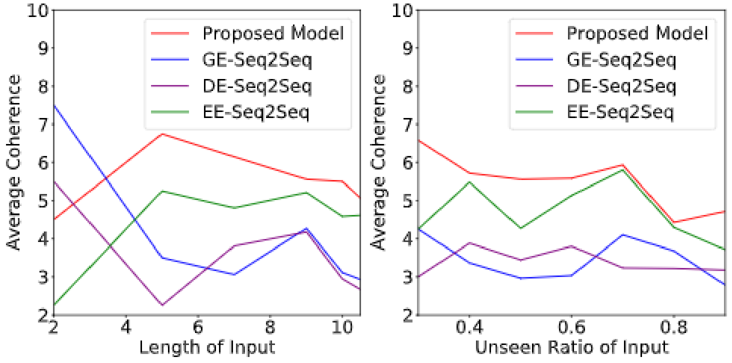


Figure 4: An illustration of how the performance is affected by the length of input (left) and the unseen ratio of input (right).

Summary

A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation(EMNLP 2018)

- Different from traditional models, the proposed model first generates a skeleton that contains the key information of a sentence, and then expands the skeleton to a complete sentence.
- The models are all scored below 6 in coherence, meaning that there is still a long way to go before the generated stories satisfy the requirement of humans.
- The above errors show that there are many dimensions in coherence, including scene-specific relevance, temporal connection, and non-recurrence.

Summary

A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation(EMNLP 2018)

- Advantage
  - The intermediate representation can better capture the relationship between the sentences.
  - The key phrase is a effective intermediate representation.
  - Jointly learning intermediate representation and story generation is a good strategy.
- Question
  - Is there a better intermediate representation.

Plan-And-Write: Towards Better Automatic Storytelling

Lili Yao[1,3*], Nanyun Peng[2*], Ralph Weischedel[2], Kevin Knight[2], Dongyan Zhao[1] and Rui Yan[1†]
liliyao@tencent.com, {npeng,weisched,knight}@isi.edu
{zhaodongyan,ruiyan}@pku.edu.cn
[1] Institute of Computer Science and Technology, Peking University
[2] Information Sciences Institute, University of Southern California, [3] Tencent AI Lab

Introduction

Plan-And-Write: Towards Better Automatic Storytelling(AAAI 2019)

- In poetry composition, Wang et al. [2016] provides a sequence of words to guide poetry generation.
- In conversational systems, Mou et al. [2016] takes keywords as the main gist of the reply to guide response generation.
- We take a similar approach to represent a story plot with a sequence of words. Specifically, we use the order that the words appear in the story to approximate a storyline.

Title (Given)	The Bike Accident
Storyline (Extracted)	Carrie $\rightarrow$ bike $\rightarrow$ sneak $\rightarrow$ nervous $\rightarrow$ leg
Story (Human Written)	Carrie had just learned how to ride a bike. She didn't have a bike of her own. Carrie would sneak rides on her sister's bike. She got nervous on a hill and crashed into a wall. The bike frame bent and Carrie got a deep gash on her leg.

Table 1: An example of title, storyline and story in our system. A storyline is represented by an ordered list of words.

Method

Plan-And-Write: Towards Better Automatic Storytelling(AAAI 2019)

- Dynamic Schema, it generates the next word in the storyline and the next sentence in the story at each step.
- Static Schema, it first generates a whole storyline which does not change during story writing.

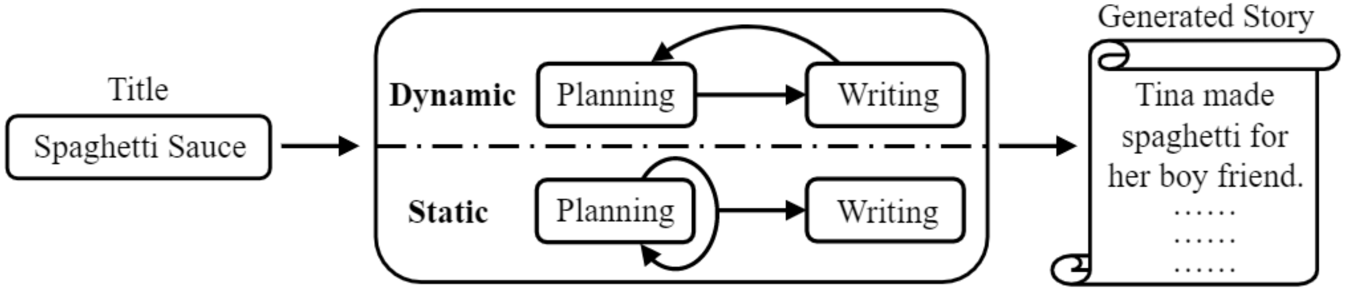
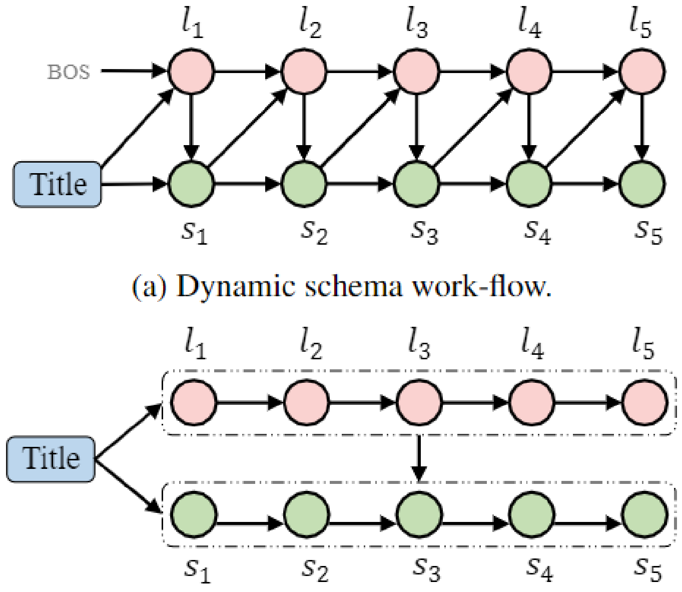


Figure 1: An overview of our system.



(a) Dynamic schema work-flow.

(b) Static schema work-flow.

Method

Plan-And-Write: Towards Better Automatic Storytelling(AAAI 2019)

- Dynamic Schema, it generates the next word in the storyline and the next sentence in the story at each step.

$$\mathcal{L}(\theta)_{dyna} = \frac{1}{N}\sum_{j=1}^{N}\left[-\log\prod_{i=1}^{m} p(s_i|\mathbf{ctx}, l_i)\right]_j$$

$$\tilde{\mathbf{h}}_{ctx} = Encode_{ctx}(\mathbf{ctx}) = [\overrightarrow{\mathbf{h}_{ctx}}; \overleftarrow{\mathbf{h}_{ctx}}]$$

$$h_y = \mathrm{GRU}(\mathrm{BOS}, C_{att}), h_w = \mathrm{GRU}(l_{i-1}, C_{att})$$

$$h_y^{'} = \tanh(W_1 h_y), h_w^{'} = \tanh(W_2 h_w)$$

$$k = \sigma(W_k[h_y^{'}; h_w^{'}])$$

$$p(l_i|\mathbf{ctx}, l_{i-1}) = g(k \circ h_y + (1-k) \circ h_w)$$

- Static Schema, it first generates a whole storyline which does not change during story writing.

$$\mathcal{L}(\theta)_{static} = \frac{1}{N}\sum_{j=1}^{N}\left[-\log\prod_{i=1}^{m} p(s_i|\tilde{\mathbf{h}}_{tl}, s_{1:i-1})\right]_j$$

$$\tilde{\mathbf{h}} = Encode(\mathbf{t}) = [\overrightarrow{\mathbf{h}}; \overleftarrow{\mathbf{h}}]$$

$$p(l_i|\mathbf{t}, l_{1:i-1}; \theta) = g(\mathrm{LSTM_{att}}(\tilde{\mathbf{h}}, l_{i-1}, \mathbf{h}_{i-1}^{\mathrm{dec}}))$$

$$\tilde{\mathbf{h}}_{tl} = Encode_{tl}([\mathbf{t}, \mathbf{l}]) = [\overrightarrow{\mathbf{h}_{tl}}; \overleftarrow{\mathbf{h}_{tl}}]$$

Experiment

Plan-And-Write: Towards Better Automatic Storytelling(AAAI 2019)

- Both dynamic and static schema significantly outperform their counterpart baseline in all evaluation aspects.

Choice %	Dynamic *vs* Inc-S2S			Static *vs* Cond-LM			Dynamic *vs* Static		
	Dyna.	Inc.	Kappa	Static	Cond.	Kappa	Dyna.	Static	Kappa
Fidelity	**35.8**	12.9	0.42	**38.5**	16.3	0.42	21.47	**38.00**	0.30
Coherence	**37.2**	28.6	0.30	**39.4**	32.3	0.35	28.27	**49.47**	0.36
Interestingness	**43.5**	26.7	0.31	**39.5**	35.7	0.42	34.40	**42.60**	0.35
Overall Popularity	**42.9**	27.0	0.34	**40.9**	34.2	0.38	30.07	**50.07**	0.38

Table 3: Human evaluation results on four aspects: fidelity, coherence, interestingness, and overall user preference. Dyna., Inc., and Cond. is the abbreviation for Dynamic schema, Inc-S2S, and Cond-LM respectively. We also calculate the Kappa coefficient to show the inter-annotator agreement.

Experiment

Plan-And-Write: Towards Better Automatic Storytelling(AAAI 2019)

- We can see that the static schema generates storylines with higher BLEU scores.
- The baselines without planning components tend to generate repetitive sentences that do not exhibit much of a story progression.
- In contrast, the plan-and-write methods can generate storylines that follow a reasonable flow, and thus help generate coherent stories with less repetition.
- The three major problems are: off-topic, repetitive, and logically inconsistent.

Method	l-B1	l-B2	l-s
Dynamic	6.46	0.79	0.88
Static	9.53	1.59	0.89

Table 6: The storyline BLEU score (only BLEU-1 and BLEU-2) and the correlation of storyline-story l-s.

Table 4: Case studies of generated storylines and stories.

Title / Problem	**Story**
Taxi / off-topic	I got a new car. It was one day. I decided to drive to the airport. I was driving for a long time. I had a great time .
Cut / repetitive	Anna was cutting her nails. She cut her finger and cut her finger. Then she cut her finger. It was bleeding! Anna had to bandage her finger.
Eight glasses/ inconsistent	Joe needed glasses. He went to the store to buy some. He did n't have any money. He found a pair that he liked. He bought them.

Summary

Plan-And-Write: Towards Better Automatic Storytelling(AAAI 2019)

- The static schema performs better than the dynamic schema because it plans the storyline holistically, thus tends to generate more coherent and relevant stories.
- We plan to extend the exploration to richer representations, such as entity, event, and relation structures, to depict story plots.
- we will explore the storyline induction and joint storyline and story generation to avoid error propagation in the current pipeline generation system. (**last paper have done!)**

Strategies for Structuring Story Generation

Angela Fan
FAIR, Paris
LORIA, Nancy
angelafan@fb.com

Mike Lewis
FAIR, Seattle
mikelewis@fb.com

Yann Dauphin
Google AI*
ynd@google.com

Introduction

Strategies for Structuring Story Generation(ACL 2019)

- Stories exhibit structure at multiple levels. While existing language models can generate stories with good local coherence, they struggle to coalesce individual phrases into coherent plots or even maintain consistency of the characters throughout the story.
- One reason for this failure is that classical language models generate the whole story at the word level, which makes it difficult to capture the high-level interactions between the plot points.

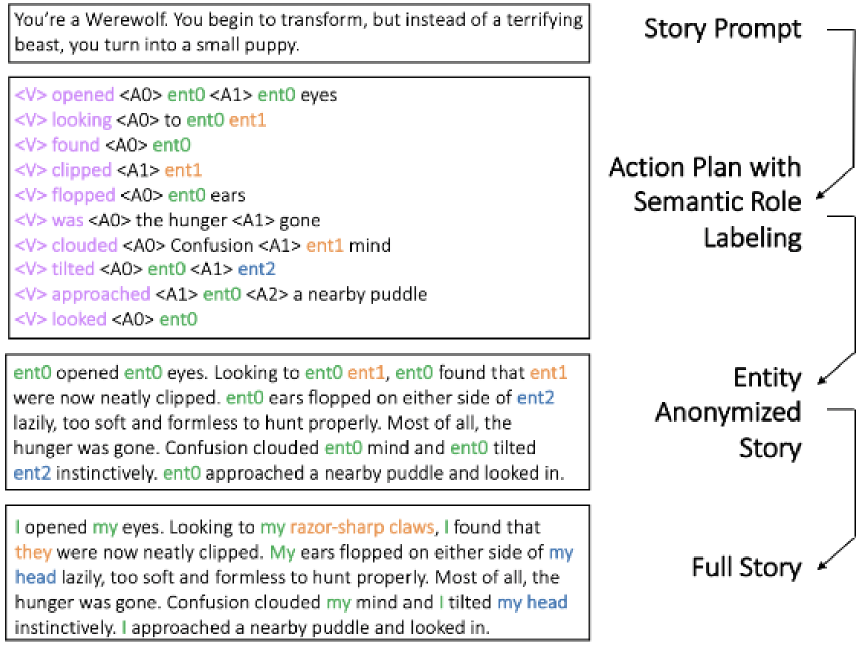


Figure 1: Proposed Model. Conditioned upon the prompt, we generate sequences of predicates and arguments. Then, a story is generated with placeholder entities such as *ent0*. Finally we replace the placeholders with specific references.

Method

Strategies for Structuring Story Generation(ACL 2019)

- Our approach breaks down the generation process in three steps: modelling the action sequence, the narrative, and then entities (such as story characters).
- To decompose a story into a structured form that emphasizes logical sequences of actions, we use Semantic Role Labeling (SRL). SRL identifies predicates and arguments in sentences, and assigns each argument a semantic role.
- To model entities, we initially generate a version of the story where different mentions of the same entity are replaced with placeholder tokens.
- Finally, we re-write these tokens into different references for the entity, based on both its previous mentions and global story context.

Experiment

Strategies for Structuring Story Generation(ACL 2019)

- Generating the SRL structure has a lower negative loglikelihood and so is much easier than generating either summaries, keywords, or compressed sentences — a benefit of its more structured form.

Decomposition	Stage 1 $-\log p(z^*)$	Stage 2 $-\log p(x\|z^*)$
Summary	4.20	5.09
Keyword	6.92	4.23
Compression	5.05	3.64
SRL Action Plan	2.72	3.95
NER Entity Anonymization	3.32	4.75
Coreference Anonymization	3.15	4.55

Table 1: Negative log likelihood of generating stories ng different decompositions (lower is better). Stage s the generation of the intermediate representation and Stage 2 is the generation of the story x condined upon z^*. Entity generation is with a word-based :abulary to be consistent with the other models.

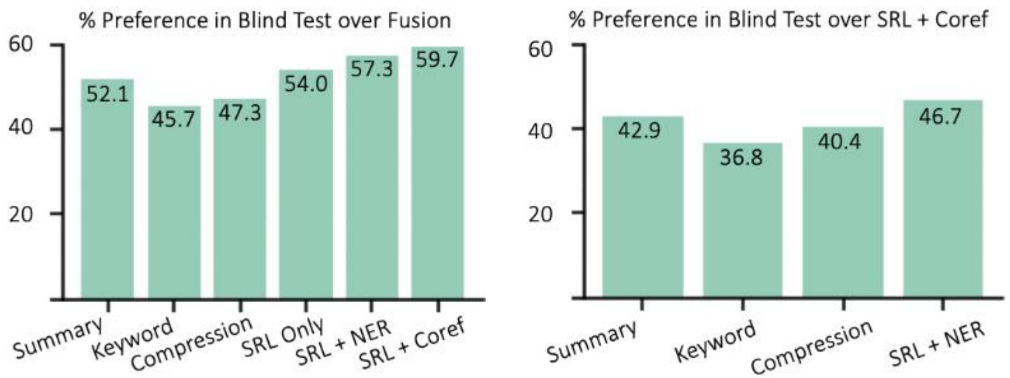


Figure 4: Human evaluations of different decomposed models for story generation. We find that using SRL action plans and coreference-resolution to build entity clusters generates stories that are preferred by human judges.

Experiment

Strategies for Structuring Story Generation(ACL 2019)

- Our decomposition using the SRL predicate argument structure improves the model's ability to generate diverse verbs.
- Adding verb attention leads to further improvement.

Model	**# Unique Verbs**	**% Diverse Verbs**
Human Stories	34.0	76.5
Fusion	10.3	61.1
Summary	12.4	60.6
Keyword	9.1	58.2
Compression	10.3	54.3
SRL	14.4	62.5
+ verb-attention	15.9	64.9

Table 2: Action Generation. Generating the SRL structure improves verb diversity and reduces repetition.

Experiment

Strategies for Structuring Story Generation(ACL 2019)

- Generated entities are often generic names such as John. In contrast, our proposed decompositions generate substantially more unique entities.
- Our full model produces more non-singleton coreference chains, suggesting greater coherence, and also gives different mentions of the same entity more diverse names.

Model	**# Unique Entities**
Human Stories	2.99
Fusion	0.47
Summary	0.67
Keyword	0.81
Compression	0.21
SRL + NER Entity Anonymization	2.16
SRL + Coreference Anonymization	1.59

Table 4: Diversity of entity names. Baseline models generate few unique entities per story. Our decompositions generate more, but still fewer than human stories. Using coreference resolution to build entity clusters reduces diversity here—partly due to re-using existing names more, and partly due to greater use of pronouns.

Model	**# Coref Chains**	**Unique Names per Chain**
Human Stories	4.77	3.41
Fusion	2.89	2.42
Summary	3.37	2.08
Keyword	2.34	1.65
Compression	2.84	2.09
SRL + NER Entity Anonymization	4.09	2.49
SRL + Coreference Anonymization	4.27	3.15

Table 5: Analysis of non-singleton coreference clusters. Baseline models generate very few different coreference chains, and repetitive mentions within clusters. Our models generate larger and more diverse clusters.

Summary

Strategies for Structuring Story Generation(ACL 2019)

- We proposed an effective method for writing short stories by separating the generation of actions and entities.
- Generating the SRL structure is much easier than generating either summaries, keywords, or compressed sentences — a benefit of its more structured form.

Summary

A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation(EMNLP 2018)

- These three paper all use Intermediate representation.
    - Keywords, key phrases, and SRL.
- The intermediate representation helps abstract the sentence and capture the relationship between the sentences better.
- A good intermediate representation, neither too simple (keywords) nor too complicated.
- In addition to the relationship between sentences in corpus, we also need to consider the relationship behind sentences. (knowledge)

Thanks