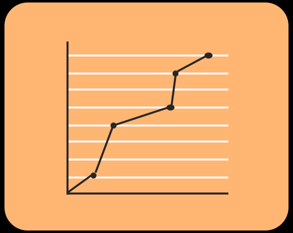
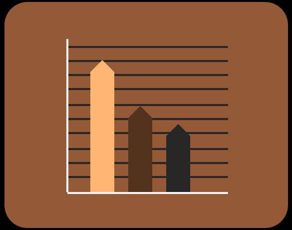




10/23/25

Controllable Neural Story Plot Generation via **Reward Shaping**

Pradyumna Tambwekar, Murtaza Dhuliawala, Lara J. Martin, Animesh Mehta, Brent Harrison, Mark O. Riedl

About The **Paper**

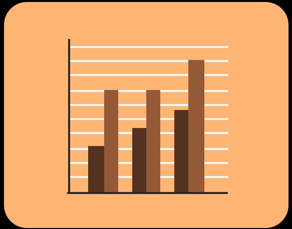
Field	**Details**
Title	Controllable Neural Story Plot Generation via Reward Shaping
Authors	Pradyumna Tambwekar, Murtaza Dhuliawala, Lara J. Martin, Animesh Mehta, Brent Harrison, Mark O. Riedl
Published in	*Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*
Place & Date	Macau, China, July 2019
Pages	5982–5988

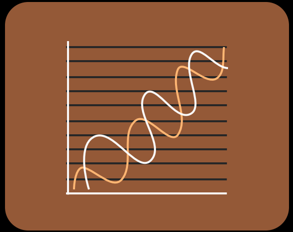
Motivation

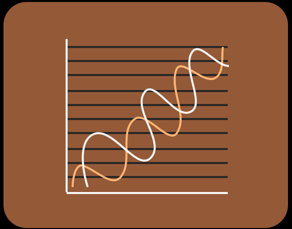
•Neural story generators often **lose direction** midway.

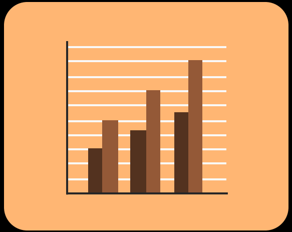
•Hard to make them **end with a specific goal** (like “two people marry”).

•Need a way to guide models without losing fluency.









Proposed methodology

- The paper introduces a **reward-shaping technique** to guide the language model toward specific goals.
- This approach ensures **controlled and coherent story progression**.

Core Idea - Reward Shaping

Train a Seq2Seq model and fine-tune it using policy gradient reinforcement learning.
Provide rewards for events that move the story closer to the goal.

Policy Gradient Update:

$$\nabla_\theta J(\theta) = R(v_{i+1})\, \nabla_\theta \log P\,(e_{i+1};\mid e_i\theta)$$

where $R(v_{i+1})$is the reward for the next verb and $P(e_{i+1};\mid e_i\theta)$is the model's predicted probability.

Event Representation

Each sentence is converted into an **event tuple**:

$$e = \langle s, v, o, m \rangle$$

where:
s= subject v = verb o = object m = modifier

Example:
"Barbara fought with Alexander." → (Barbara, fought, Alexander)

Reward Function

Dense rewards encourage gradual progress toward the goal.

- **Distance Reward:**

$$r_1(v) = \log \sum_{s \in S_{\mathrm{v,g}}} \left(l_s - d_s(v, g)\right)$$

- **Frequency Reward:**

$$r_2(v) = \log \frac{k_{v,g}}{N_v}$$

- **Final Reward:**

$$R(v) = \alpha \times r_1(v) \times r_2(v)$$

Rewards verbs that appear close to and frequently before the goal verb.

Verb Clustering

- Verbs are grouped using **Jenks Natural Breaks** based on their reward values.
- The model's output verb is limited to the next cluster, ensuring smoother narrative progression.
- Prevents the model from jumping directly to the target event.

Model Setup

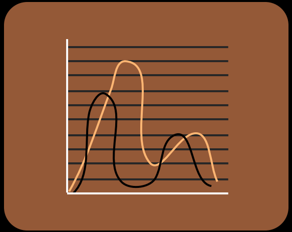
Model	Description
Seq2Seq	Baseline event2event model
DRL-Unrestricted	RL without verb clustering
DRL-Clustered	RL with reward shaping and verb clustering

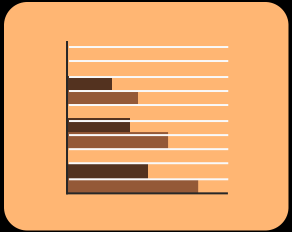
Encoder–Decoder LSTM (1024 hidden units)

Batch size: 64 Pretraining: 200 epochs RL fine-tuning: 200 epochs

Dataset

- **CMU Movie Summary Corpus (Bamman et al., 2013)**
- Clustered into 100 genres using Latent Dirichlet Allocation (LDA).
- Selected one cluster containing *soap-opera–like* stories.
- Target verbs: *admire* and *marry* .
- 90% training data and 10% testing data.
- Focus: Romance-style plots and relationship-driven stories.





Findings:

- DRL-clustered achieves **>93% goal completion**, lowest perplexity and concise story length. proving **reward shaping & clustering work**.
- Perplexity drops drastically vs. baseline → model better matches corpus distribution.
- Slightly shorter stories → **faster goal achievement without skipping coherence**.
- **Perplexity Formula:**

$$\text{Perplexity} = e^{-\frac{1}{N}\sum_i \log P(x_i)}$$

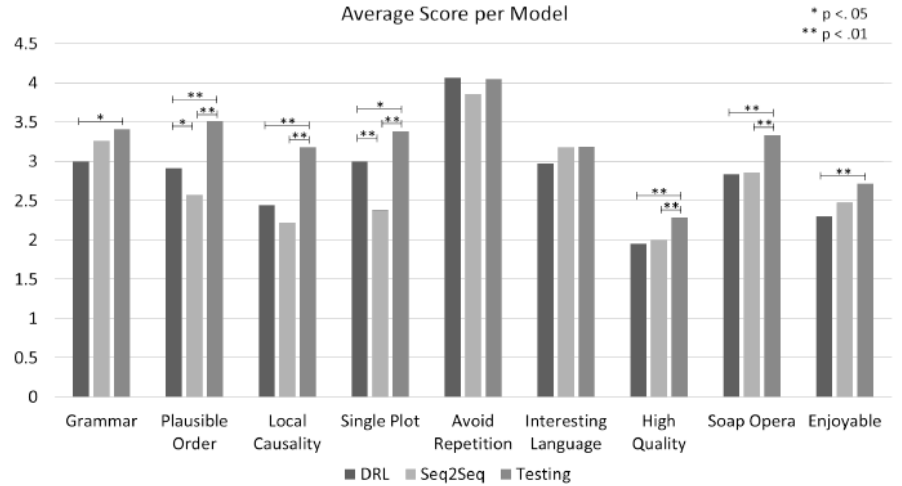
Goal	Model	Goal achievement rate	Average perplexity	Average story length
admire	Test Corpus	20.30%	n/a	7.59
	Seq2Seq	35.52%	48.06	7.11
	Unrestricted	15.82%	**5.73**	7.32
	Clustered	**94.29%**	7.61	4.90
marry	Test Corpus	24.64%	n/a	7.37
	Seq2Seq	39.92%	48.06	6.94
	Unrestricted	24.05%	9.78	7.38
	Clustered	**93.35%**	**7.05**	5.76

Human Evaluation & Conclusions

Human Evaluation

- 26 translators converted event sequences into readable sentences.
- 175 participants on Amazon Mechanical Turk rated 9 aspects (1–5 scale): grammar, order, coherence, enjoyment, etc.
- Compared results for Testing Corpus, Seq2Seq and DRL models.
- Analysis used **one-way ANOVA** with **Tukey HSD** post-test.

Table 2: An example eventified story from the DRL-clustered system paired with the translation written by a pair of participants.



Key Findings

DRL model scored highest for:

- Plausible Order ($p < 0.05$)
- Single Plot Coherence ($p < 0.05$)
- Overall Quality ($p < 0.01$)

No degradation in grammar, repetition, or enjoyment.

Testing corpus rated highest for *Soap Opera* style (genre fidelity).

Strengths and Weaknesses

Strengths:

- Strong results
- Good Human Evaluation

Weaknesses:

- Limited Control
- Narrow Testing

Resources

- Tambwekar, Pradyumna, et al. "Controllable neural story plot generation via reward shaping." *arXiv preprint arXiv:1809.10736* (2018).
- Martin, Lara, et al. "Event representations for automated story generation with deep neural nets." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. No. 1. 2018.

THANKS!

DO YOU HAVE ANY QUESTIONS?