Modeling Protagonist Em🙃tions for Emotion-Aware Storytelling

Faeze Brahman

Snigdha Chaturvedi

Motivation

What makes stories interesting and engaging to readers?

- ⭐ The emotion the storyteller puts into the narrative by connecting moments with characters, their motivations, goals, and achievements.

Modeling the
emotional trajectory
of the protagonist

Overview

Task: Emotion-Aware Storytelling

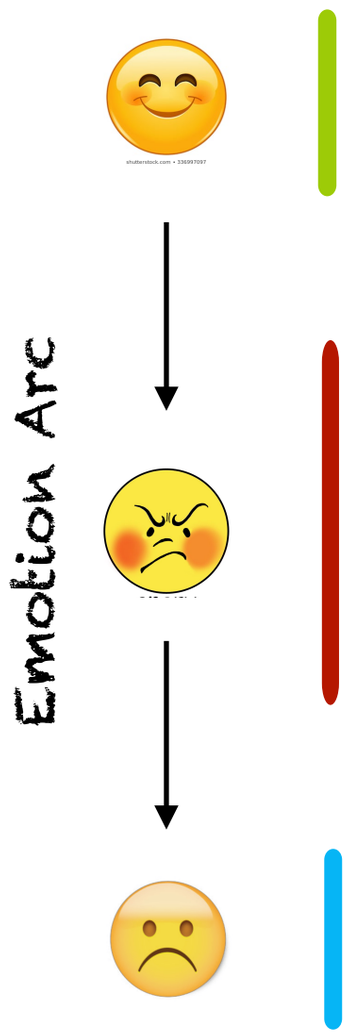
Models: Emotion-Reinforced (EmoRL) models

Dataset : Annotation Pipeline

Evaluation: Automatic / Human

Task Definition

Tell me a story about **Raw Burger:**



Tom went to a burger place with his friends.

Protagonist

He ordered a burger. When he got it , **he** noticed that it was raw. **Tom** yelled at the waiter for it being raw.

He was really disappointed.

What if we want to generate stories with different emotion arcs?

Title: Dance

fear - joy - joy

Kelly was worried about her dance recital. She had practiced her dance for weeks. She decided to try out for the school's dance team. Kelly was nervous but knew she could do well. She was so excited she gave her best impression!

Sadness - joy - joy

I was very depressed. I went to a dance class with a friend of mine. We tried out some different moves. We got stuck dancing for a long time. The next day I tried out some new moves and got a standing ovation.

Challenges

- It is difficult to balance fluency and expressions of emotions in a natural and coherent way.
- High-quality large-scale emotion-annotated data are hard to obtain.

Automatically annotate story corpus using COMET (Bosselut et al. 2019)

Overview

Task: Emotion-Aware Storytelling

Models: Emotion-Reinforced (EmoRL) models

Dataset : Annotation Pipeline

Evaluation: Automatic / Human

Overview

Task: Emotion-Aware Storytelling

Models: Emotion-Reinforced (EmoRL) models

Dataset : Annotation Pipeline

Evaluation: Automatic / Human

Approach

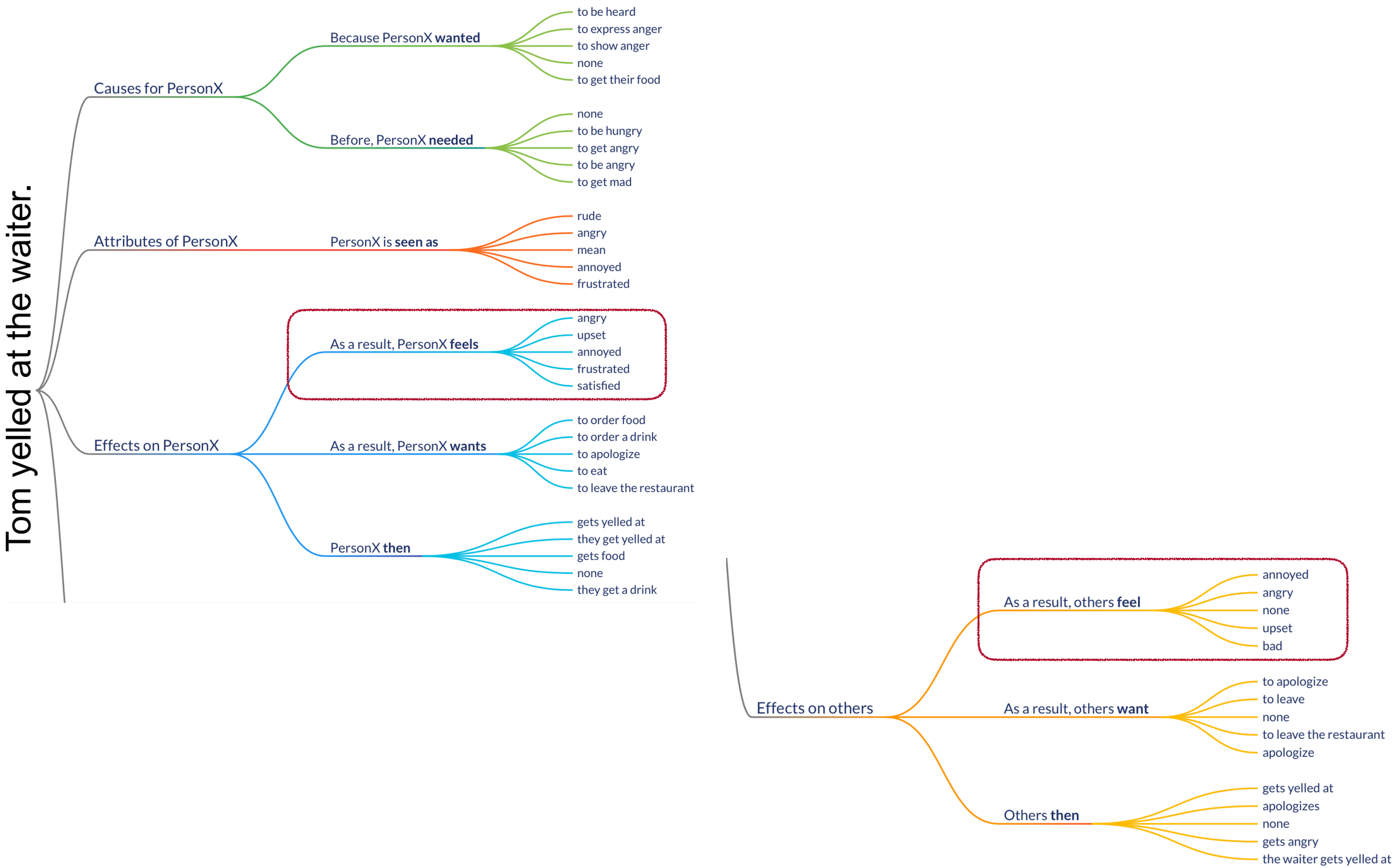
We build our model based on GPT2 (Radford et al. 2019):

We design two special rewards to *regularize the story generation* and *control emotion arc of the protagonist* through Reinforcement Learning.

We explore two different ways of tracking the emotions:

- 🤖 Commonsense Transformer (COMET)
- 🤖 Emotion Classifier

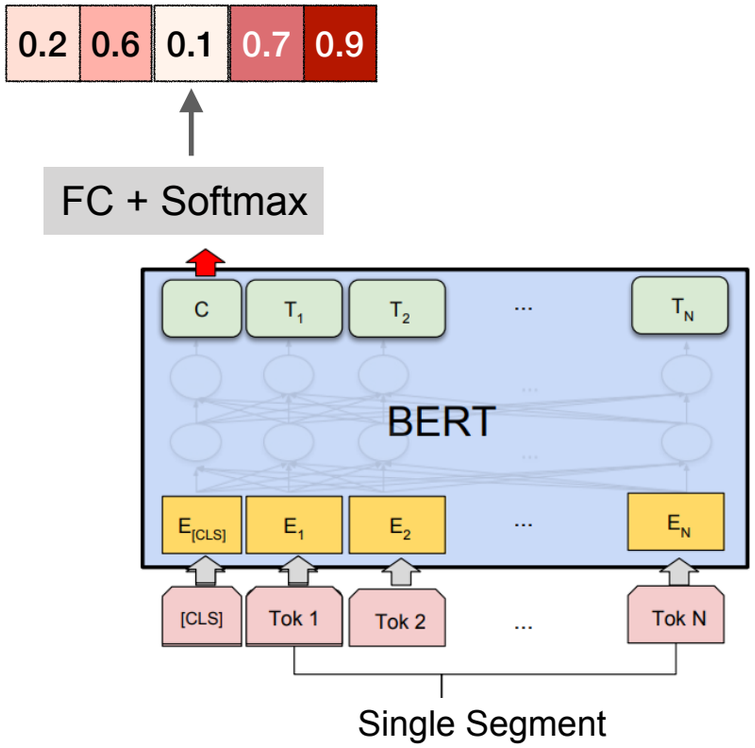
Tracking Emotion — COMET



Faeze Brahman and Snigdha Chaturvedi . Modeling Protagonist Emotions for Emotion-Aware Storytelling . EMNLP 2020

Tracking Emotion — Classifier

Adapt a pre-trained BERT for multi-label classification over 5 basic emotions



Two-step fine-tuning:

- On human-annotated dataset (tweets)
- Further fine-tune on story training data — automatically annotated with the protagonist's emotions

Dataset: Human-annotated Tweets + Auto-annotated Stories

Storytelling Model Arch.

The story generator interact with environment during training to regularize the generation

- Emotion Matching (Ec-Em)
- Emotion Classifier (Ec-Clf)

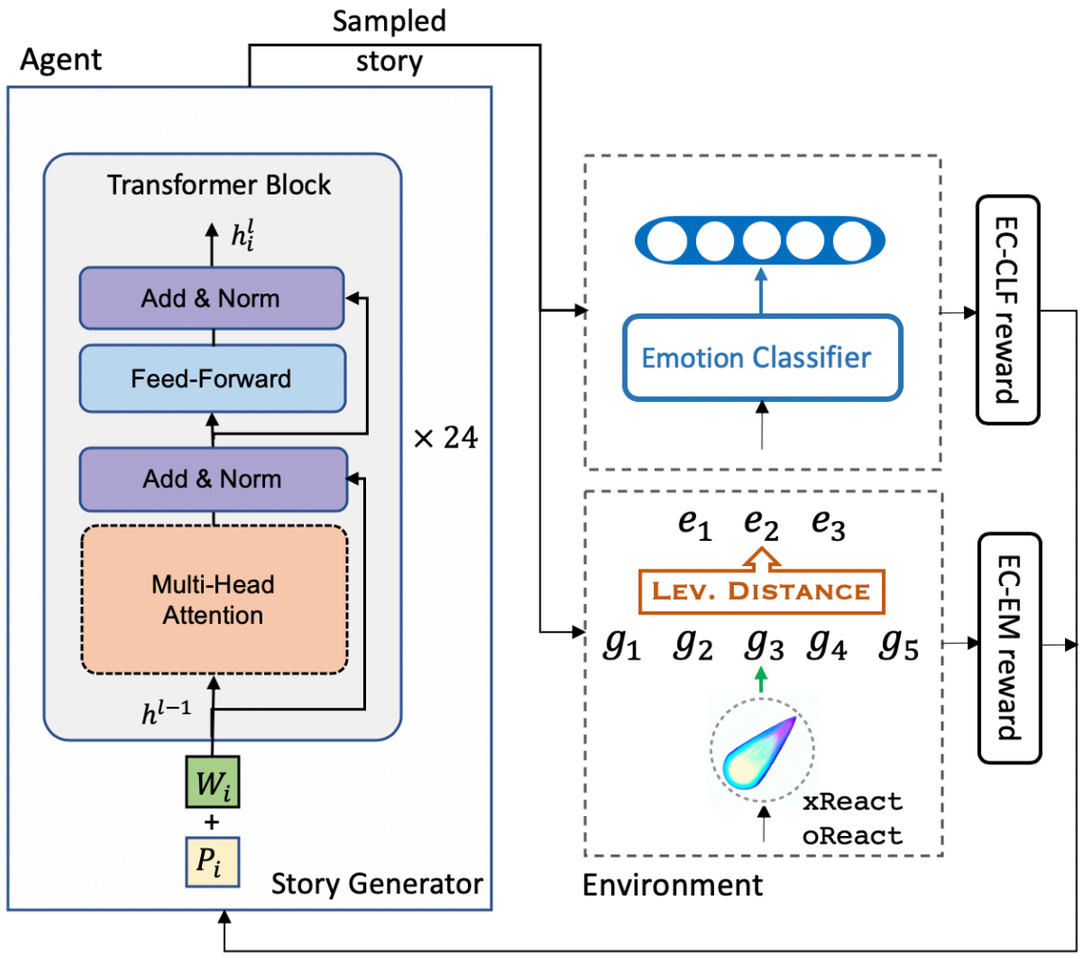
Input:

Title $t = \{t_1, t_2, \ldots, t_m\}$

Emotion arc $a^* = \{e_1, e_2, e_3\}$

Output:

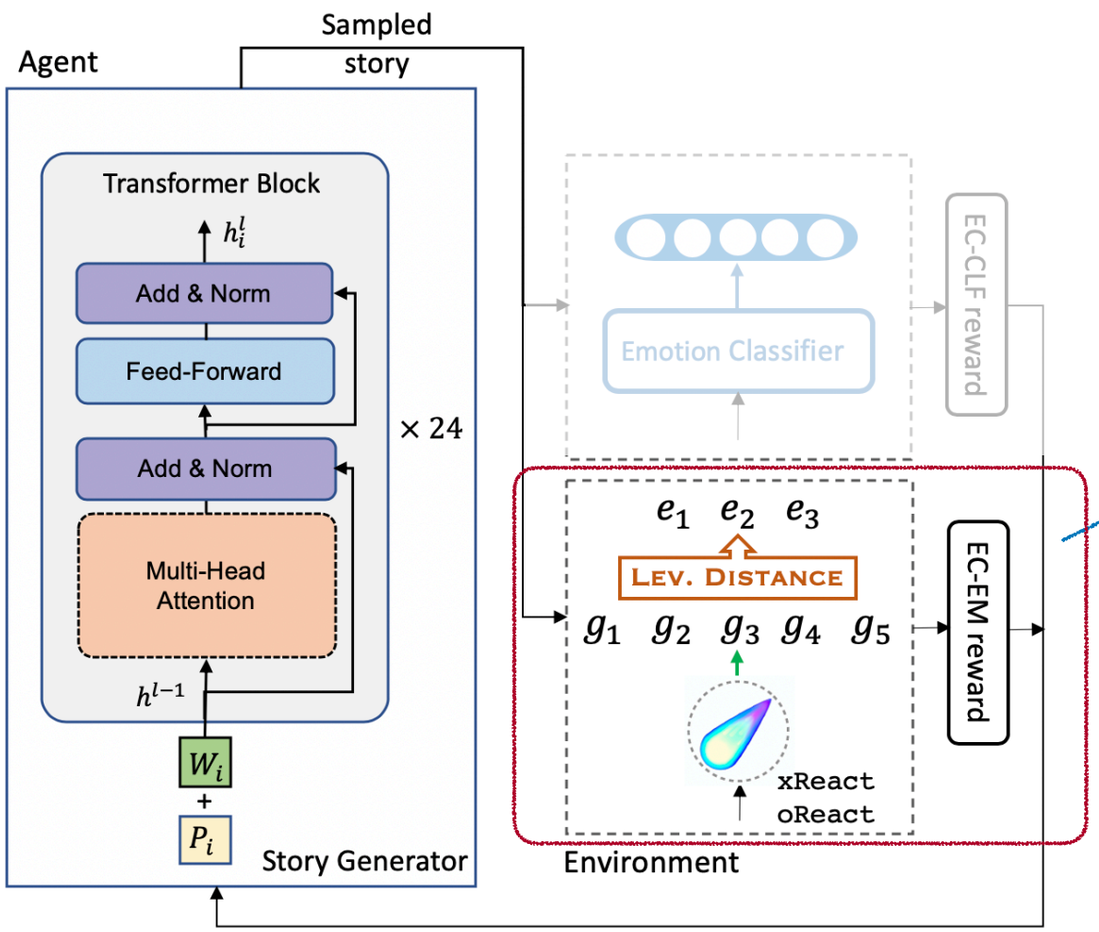
Story $Y = \{y_1, y_2, \ldots, y_n\}$



GPT2-M

Faeze Brahman and Snigdha Chaturvedi . Modeling Protagonist Emotions for Emotion-Aware Storytelling . EMNLP 2020

EmoRL — RL-EM



GPT2-M

Ec-Em reward quantify the string alignment of the emotion arc of the generated story to desired emotion arc.

Generated story emotional reactions:

$$a^g = \{g_1, g_2, \ldots, g_N\}$$

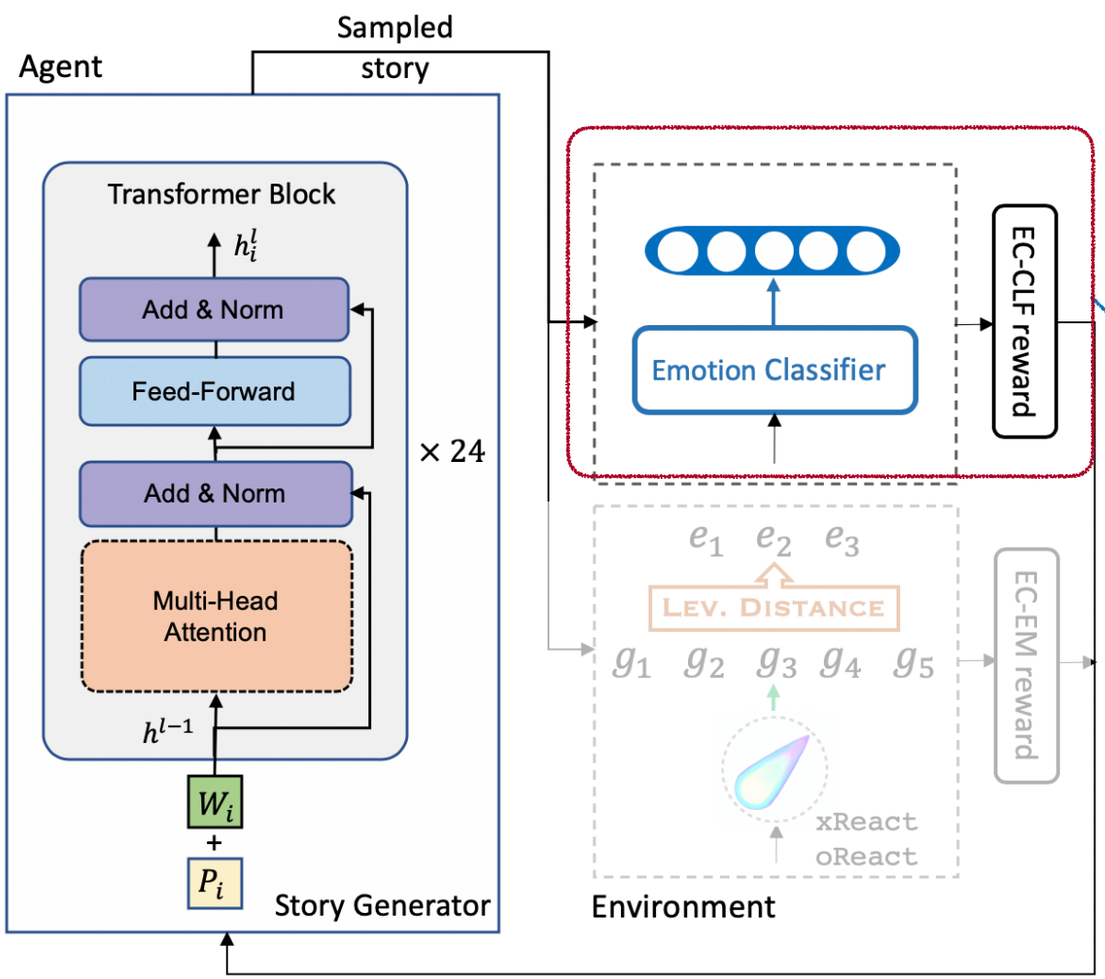
Desired Emotion arc:

$$a^* = \{e_1, e_2, e_3\}$$

Reward:

$$r_{em} = \text{levenshtein}(a^g, a^*)$$

EmoRL — RL-CLF



GPT2-M

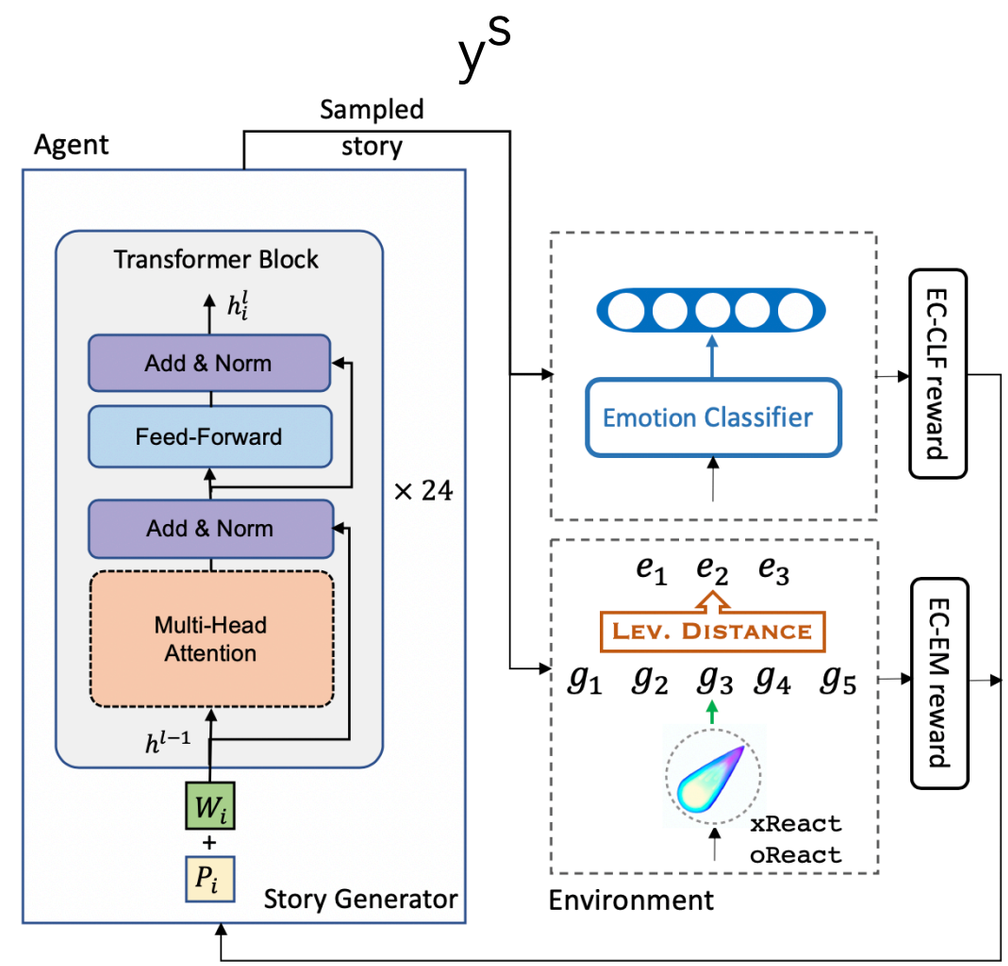
Reward:

$$r_{clf} = 1/k \sum_{j=1}^{k} p_{clf}(e_j^* | x_j)$$

Ec-Clf reward measures the probability of desired emotion arc using emotion classifier.

Policy Gradient

self-critical: $$\mathcal{L}_{RL} = -(r(\boldsymbol{y}^s) - r(\hat{\boldsymbol{y}})) \sum_{i=k+m}^{k+m+n} \log p_\theta(y_i^s | y_{<i}^s)$$



GPT2-M

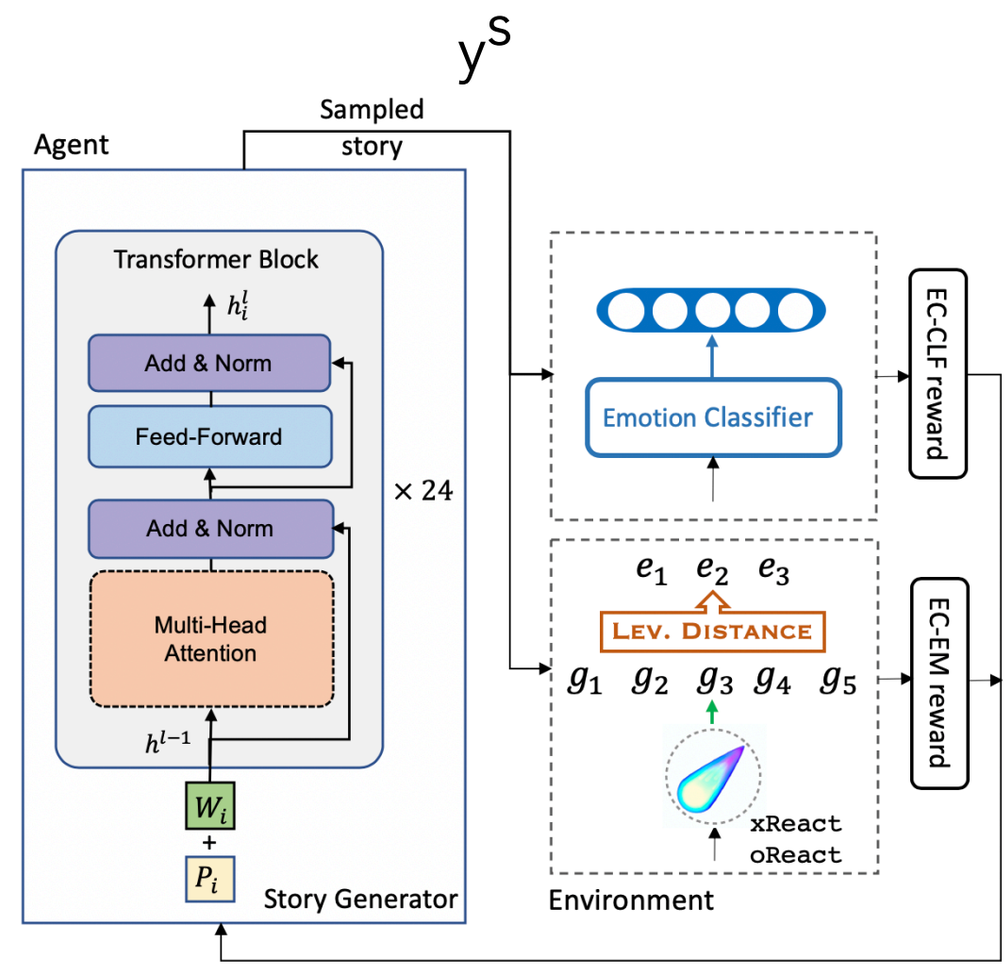
Faeze Brahman and Snigdha Chaturvedi . Modeling Protagonist Emotions for Emotion-Aware Storytelling . EMNLP 2020

Policy Gradient

self-critical: $$\mathcal{L}_{RL} = -(r(\boldsymbol{y}^s) - r(\hat{\boldsymbol{y}})) \sum_{i=k+m}^{k+m+n} \log p_\theta(y_i^s | y_{<i}^s)$$

ML loss: $$\mathcal{L}_{ML} = -\sum_{i=m}^{m+n} \log p(y_i | y_{<i}, \boldsymbol{t})$$

Mixed loss: $$\mathcal{L}_{mixed} = \gamma \mathcal{L}_{RL} + (1-\gamma)\mathcal{L}_{ML}$$



GPT2-M

Faeze Brahman and Snigdha Chaturvedi . Modeling Protagonist Emotions for Emotion-Aware Storytelling . EMNLP 2020

Overview

Task: Emotion-Aware Storytelling

Models: Emotion-Reinforced (EmoRL) models

Dataset : Annotation Pipeline

Evaluation: Automatic / Human

Overview

Task: Emotion-Aware Storytelling

Models: Emotion-Reinforced (EmoRL) models

Dataset : Annotation Pipeline

Evaluation: Automatic / Human

Dataset — Annotation Pipeline

ROCStories Corpus
(Mosatafazadeh et al.)

Iris hired a babysitter for her daughter.

She planned on going on a date with her husband.

Iris waited for the sitter to show.

Ultimately, the sitter never showed up.

As a result, Iris had to stay home.

(1) Coref Resolution

Character mentions

Dataset — Annotation Pipeline

ROCStories Corpus
(Mosatafazadeh et al.)

Iris hired a babysitter for **her** daughter.

She planned on going on a date with **her** husband.

Iris waited for the sitter to show.

Ultimately, the sitter never showed up.

As a result, **Iris** had to stay home.

(2) Protag. Identification

Character with the most mentions:
Iris

Dataset — Annotation Pipeline

ROCStories Corpus
(Mosatafazadeh et al.)

Agent
Iris hired a babysitter for her daughter.

Agent
She planned on going on a date with her husband.

Agent
Iris waited for the sitter to show.

Other
Ultimately, the sitter never showed up.

Agent
As a result, Iris had to stay home.

(3) Protag. Role Identification

Protagonist's roles:
Agent / Other

Dataset — Annotation Pipeline

ROCStories Corpus
(Mosatafazadeh et al.)

Iris hired a babysitter for her daughter.	Satisfied
She planned on going on a date with her husband.	Anxious
Iris waited for the sitter to show.	Anxious
Ultimately, the sitter never showed up.	Annoyed
As a result, Iris had to stay home.	Sad

(4) Generate Emotion Reaction

Protagonist's Emotional reactions

Dataset — Annotation Pipeline

ROCStories Corpus
(Mosatafazadeh et al.)

Iris hired a babysitter for her daughter.

She planned on going on a date with her husband.

Iris waited for the sitter to show.

Ultimately, the sitter never showed up.

As a result, Iris had to stay home.

Satisfied

Annoyed

Sad

(5) Segmentation

Dataset — Annotation Pipeline

ROCStories Corpus
(Mosatafazadeh et al.)

Iris hired a babysitter for her daughter.

She planned on going on a date with her husband.

Iris waited for the sitter to show.

Ultimately, the sitter never showed up.

As a result, Iris had to stay home.

joy

~~Satisfied~~

anger

~~Annoyed~~

sadness

~~Sad~~

(6) Map to Basic Emo via NRC

Emotion arc

Overview

Task: Emotion-Aware Storytelling

Models: Emotion-Reinforced (EmoRL) models

Dataset : Annotation Pipeline

Evaluation: Automatic / Human

Overview

Task: Emotion-Aware Storytelling

Models: Emotion-Reinforced (EmoRL) models

Dataset : Annotation Pipeline

Evaluation: Automatic / Human

Baselines

We use the following baselines:

- GPT2 + FT

Not provided emotion

- Fusion + Emo (Fan et al 2018)
- Plan&Write + Emo (Yao et al. 2019)
- PPLM (Dathathri et al. 2020)
- EmoSup (GPT + FT + Emo)

Automatic Evaluation

Content quality:

- PPL, BLEU, Distinct-n, repetition-4

Emotion faithfulness:

- Seg-word & Arc-word: percent of segments/stories that contain corresponding emotion words.
- Seg-acc & Arc-acc: emotion accuracy of generated segments/stories as identified by emotion classifier.
- EC-CLF & EC-EM (RL rewards)

Automatic Evaluation

Models	PPL (↓)	BLEU-1 (↑)	BLEU-2 (↑)	Dist-1 (↑)	Dist-2 (↑)	Dist-3 (↑)	Repet-4 (↓)
Fusion + Emo	24.02	21.10	2.61	66.18	90.88	96.91	23.30
Plan&Write + Emo	17.43	22.46	3.03	66.32	90.47	95.59	28.61
PPLM-3	–	20.36	2.37	71.37	93.90	98.19	13.36
PPLM-5	–	20.61	2.47	71.47	93.99	98.21	14.02
GPT-2 + FT*	12.16	22.68	3.10	72.93	94.24	98.28	12.10
EmoSup	**11.10**	22.70	3.23	71.44	93.75	98.10	13.94
RL-EM	11.98	22.52	3.15	**73.32**	**94.76**	**98.56**	**10.09**
RL-CLF	11.31	**22.78**	**3.26**	71.16	93.65	98.05	13.34

Outperform GPT2+FT on PPL

Best PPL by EmoSup

Best BLEU by RL-Clf

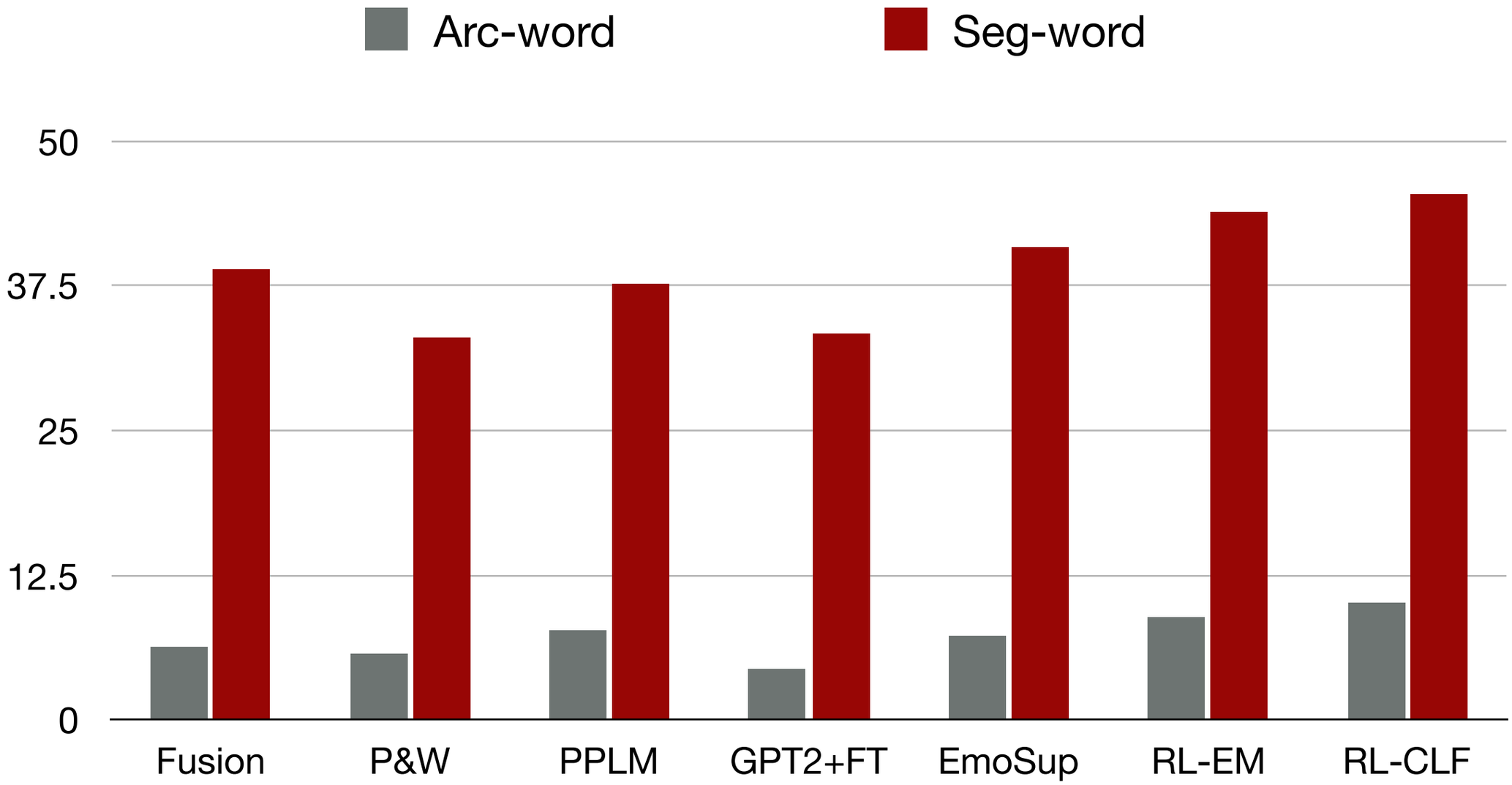
Best diversity by RL-Em

- Both Emo-reinforced models outperform all baselines for BLEU and diversity/repetition scores respectively.

Automatic Evaluation

Emotion Faithfulness

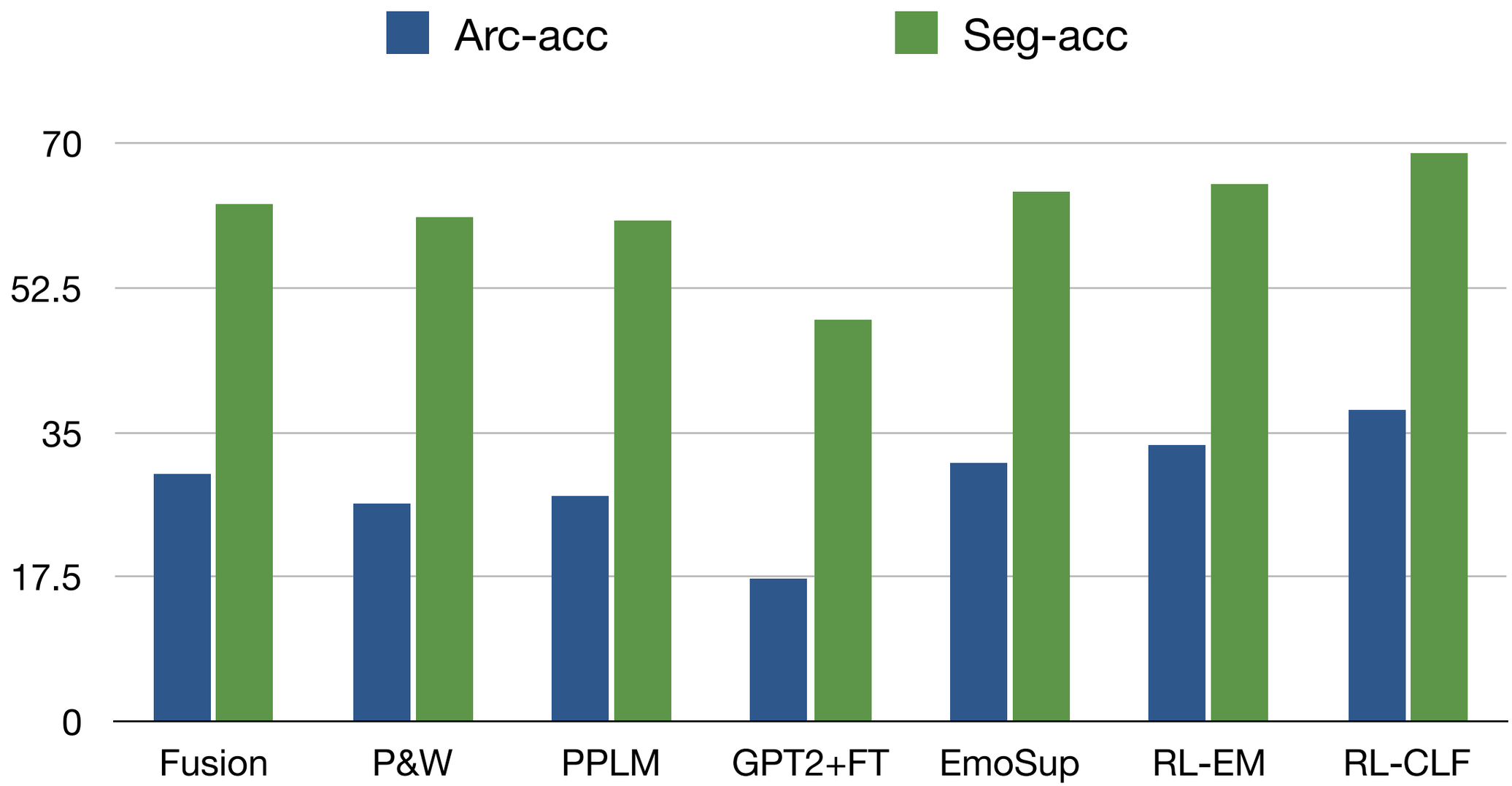
RL-Clf achieves the best performance followed by RL-Em



Automatic Evaluation

Emotion Faithfulness

All models are superior to GPT2+FT (as expected). RL-Clf is the best.



Human Evaluation

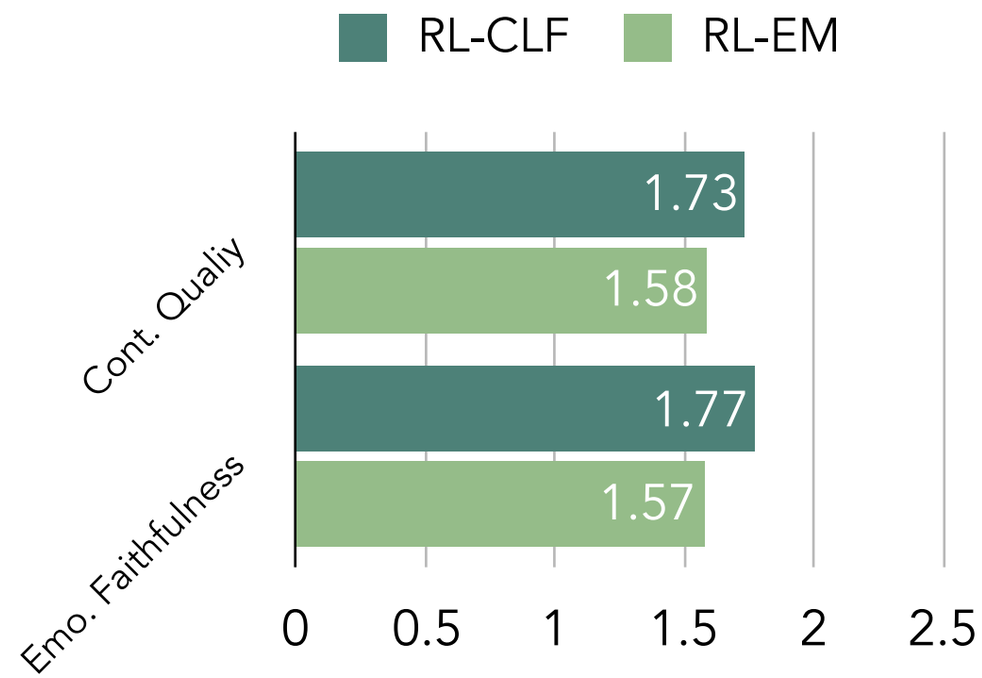
Test on 100 randomly selected title + emotion arc

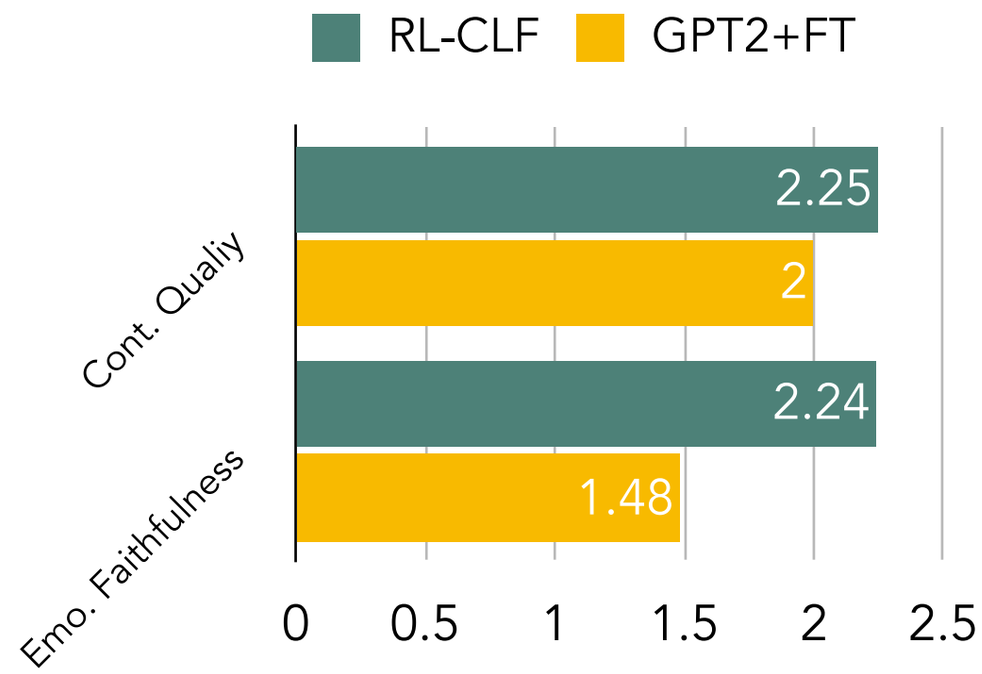
Evaluation criteria:

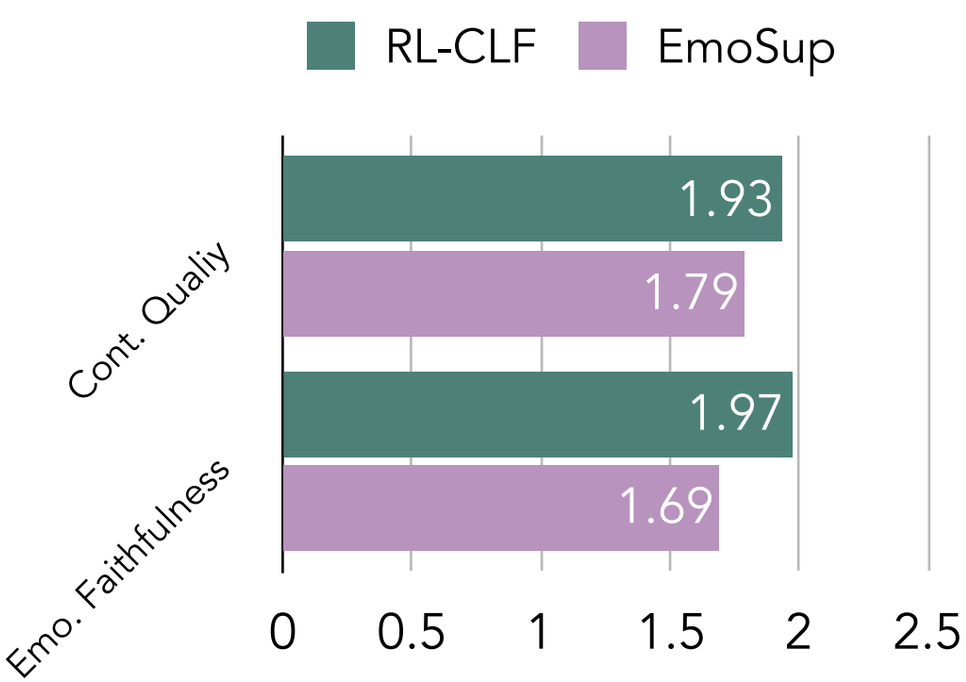
- Emotion faithfulness (likert scale 0-3)
- Content quality (likert scale 0-3)
- Overall preference

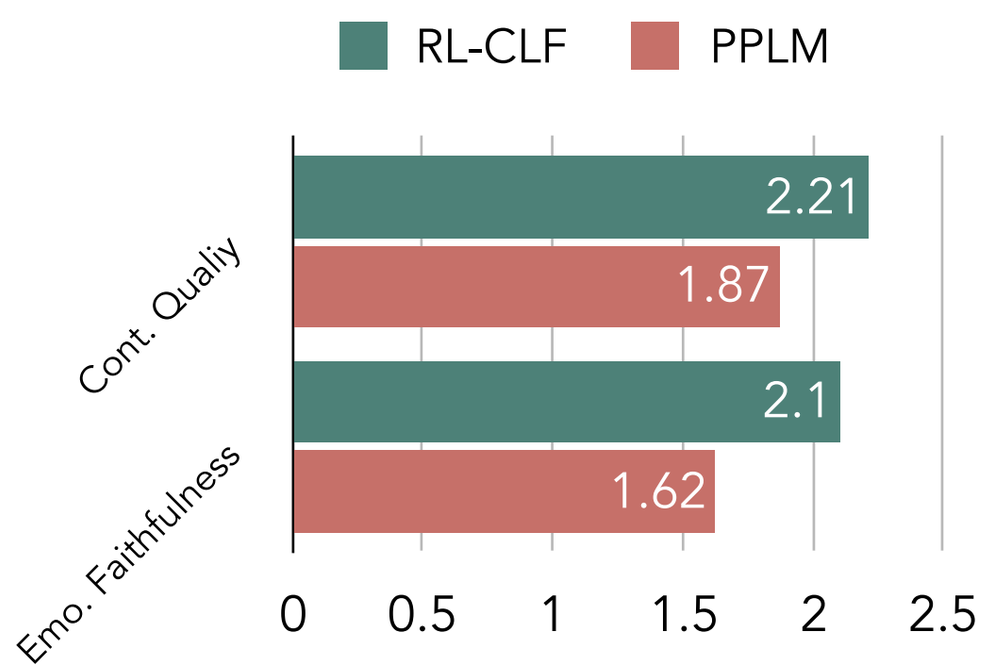
RL-Clf improves over RL-Em on Emo. and Qual.

RL-Clf is not only better at adhering to emotion arc (+0.76) but also generates better content (+0.25)





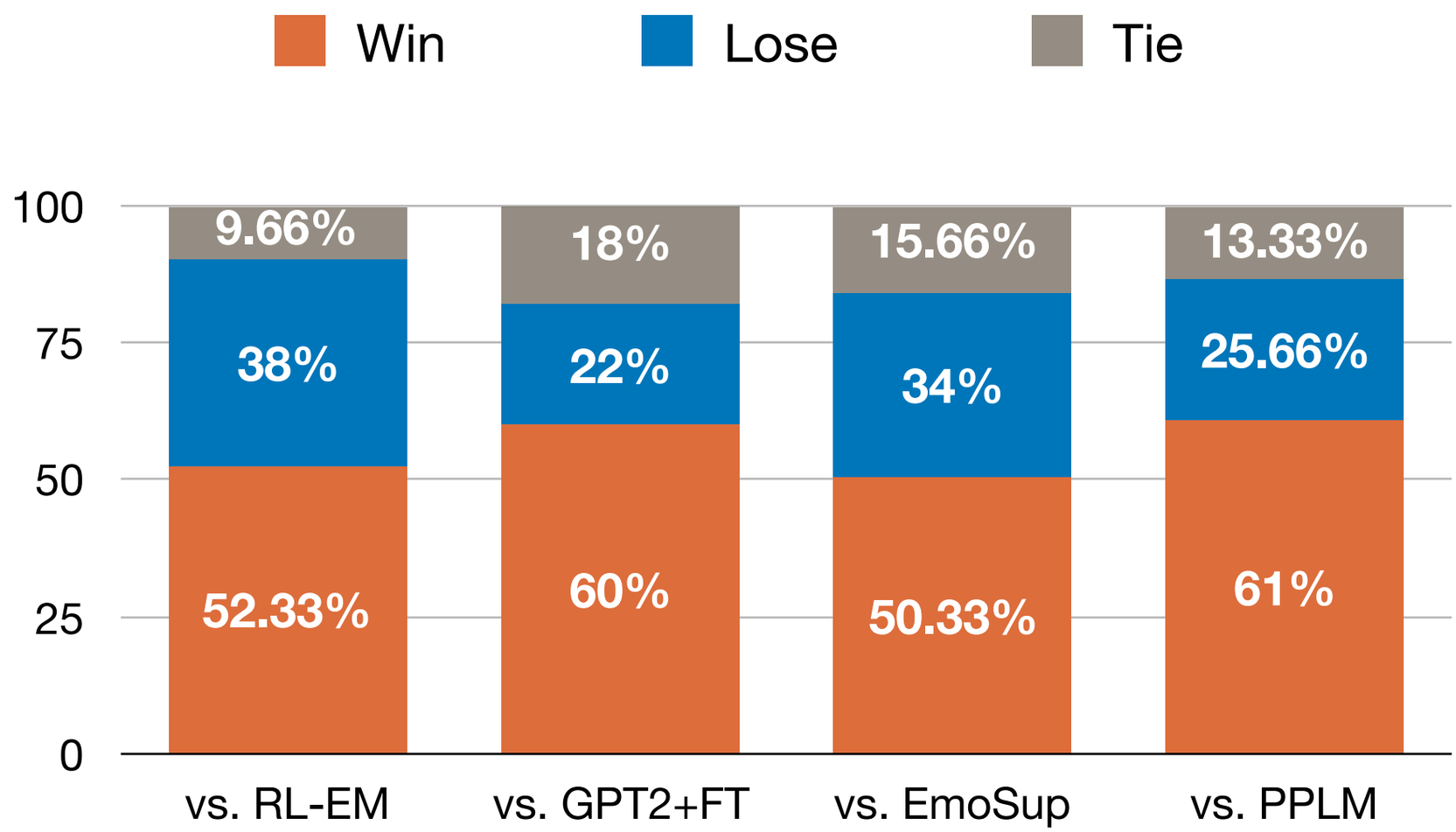




Human Evaluation Results

Overall Preference of RL-CLF

RL-Clf is preferred over other models!



Summary

- ⭐ We present the first study on modeling the *emotion arc* of the protagonist in neural story generation.
- ⭐ We propose two emotion-consistency rewards designed to enforce the desired *emotion arcs* using RL.
- ⭐ We track the protagonist's emotions using (1) commonsense knowledge models, and (2) emotion classifier trained using transfer learning.
- ⭐ We empirically demonstrate that our models can effectively generate stories that follow the desired *emotion arc* w/o sacrificing the story quality.

Thank you!
Questions?