
Controlled Affective Text Generation

Ishika Singh* Ahsan Barkati* Tushar Goswamy Ashutosh Modi
Indian Institute of Technology Kanpur (IITK)
{ishikas, ahsanb, tgoswamy}@iitk.ac.in
ashutoshm@cse.iitk.ac.in

Abstract

Humans use language not just to convey information but also to express their implicit feelings and mental states. In this work, we propose an affective (emotional) text generation model based on the SOTA language generation models (e.g., GPT-2). We posit a model capable of generating affect-driven and topic focused sentences without losing grammatical correctness as the affect intensity increases. The model will give the user the flexibility to control the category and intensity of emotion as well as the subject of the generated text. Previous attempts at modelling fine-grained emotions fall out on grammatical correctness at extreme intensities, but our model is resilient to this and delivers robust results at all intensities. We conduct human evaluations in which our model outperforms existing affective text generation models. The code is made available at this repository: <https://github.com/ishikasingh/Affective-text-gen>

1 Introduction

Emotion in language plays a critical role in conveying mental and emotional states, along with the information intended to be conveyed. There needs to be automated processing and generation of affect-driven language, to improve the quality of conversations generated and make language models more empathetic towards a human user [Colombo et al., 2019]. Current research on incorporating language in robot learning [Bisk et al., 2020, Luketina et al., 2019] can potentially be benefited not just by enabling affective human-robot dialog, but also in optimizing other reinforcement learning components. For instance, user’s responses can be exploited to extract implicit emotion-based reward cues [Sumers et al., 2020], and the robot can respond in contrasting emotion to gain user’s trust. In interactive narrative generation [Ammanabrolu et al., 2020], an affect-incorporated language model will improve user’s experience by making a long-lasting impact [Mar et al., 2011].

In particular, we propose coarse and fine-grained affective text generation model, built on top of GPT-2 [Radford et al., 2019]. Our model provides interactive degrees of freedom in terms of the choice of the emotion category (ranging over 8 basic emotions - joy, sadness, anger, fear, trust, surprise, anticipation, disgust), with fine-grained control over emotion intensity for each category, and the topic of the generated text. We provide model evaluations to justify that the quality of text generated by our model both in terms of perceived emotion and grammatical correctness is considerably better than the existing system (AffectLM [Ghosh et al., 2017]).

2 Related Works

In recent times, neural models for emotional text generation have been proposed. Affect-LM Ghosh et al. [2017] uses an LSTM-based approach for generating expressive emotional text. It is capable of generating sentences in 4 affect categories (Positive, Anxious, Sadness and Anger), and the affect intensity can be varied on a scale of 0 to ∞ . However, since its introduction, several new text

* The two authors contributed equally to this work.

generating language models have been proposed (e.g., GPT-2 Radford et al. [2019]) which have outperformed previous RNN based language generation models. The Affect-LM model depreciates in the grammatical correctness of its generated sentences as the affect intensity is increased to the higher end of the spectrum, and is limited to 4 affect categories. The Plug and Play Language Models (PPLM) Dathathri et al. [2020] combines a pre-trained language model like GPT-2 with attribute classifiers that guide text generation. It enables the user to control the topic, sentiment (positive/negative) and the strength of the influence of these attributes (using the stepsize of a gradient descent equation) for the generated sentences. PPLM model also fails to generate grammatical text, when emotion intensity is increased, and is limited to 2 sentiment categories.

3 Method

Our model is based on the GPT-2 [Radford et al., 2019] text generation model. Inspired by the approach in Plug and Play Language Model (PPLM) [Dathathri et al., 2020], we inject an affect attribute probability based perturbation in the base probability distribution. GPT-2 is a transformer-based sentence-generative language model. It maintains the context as a history embedding matrix H_t , and recursively samples next word using current token embeddings O_{t+1} obtained using the context. As in PPLM, we perturb the GPT-2 model in the following manner,

$$H'_t = H_t - \eta \frac{\partial \text{Loss}}{\partial H_t}; O'_{t+1}, H_{t+1} = \text{LM}(s_t, H'_t); s_{t+1} \sim p'_{t+1} = \text{Softmax}(W \cdot O'_{t+1}) \quad (1)$$

where H_t is the history embedding matrix, O_{t+1} is the token embedding used to get the next word's probability distribution p_{t+1} by learning a parameter W , and s_{t+1} represents the sampled word at $(t + 1)$ th iteration. H'_t, O'_{t+1}, p'_{t+1} are the perturbed representations. The Loss in Eq.1 is defined as,

$$\text{Loss} = \text{KLD}_{\text{perturbed-unperturbed}} + \text{Loss}_{\text{topic}} + \alpha \text{Loss}_{\text{affect}} \quad (2)$$

$$\text{Loss}_{\text{topic}} = -\log(\sum(\text{BoW probs})); \text{Loss}_{\text{affect}} = -\log((\text{BoW probs}) \cdot \mathcal{N}(\text{affectInt}, \text{knob}, \text{var})) \quad (3)$$

In Eq.2, the Loss consists of a KL-Divergence term, which keeps the perturbed next word probability distribution close to the actual one, hence ensuring grammatical correctness. The second term corresponds to topic attribute and the third corresponds to affect attribute. In $\text{Loss}_{\text{affect}}$, $\mathcal{N}(\text{affectInt}, \text{knob}, \text{var})$ is a Gaussian function, which controls the intensity of the affective text generated. Here 'affectInt' represents the intensity values for the BoWs corresponding to a emotion category, ranging from 0 to 1, where 1 is the maximum intensity, knob (the mean of the Gaussian) scales-up the values for the words closer to it, hence increasing the probability of the words with intensity values as defined by the knob , and var provides with flexibility on the intensity range to be picked as the next word. To incorporate affect, we use human-annotated affect information provided by NRC Emotion Intensity Lexicon [Mohammad, 2018] with 10,000 words belonging to eight basic emotions, which fulfils both BoWs emotion categories and emotion intensities.

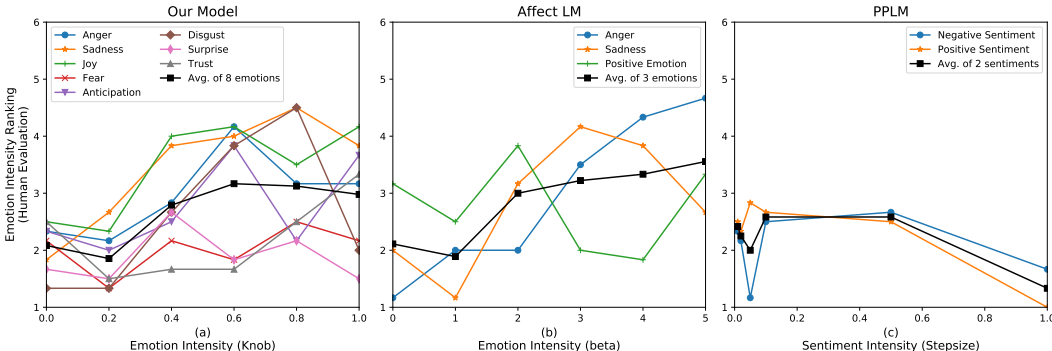


Figure 1: Human Perceived Intensity Evaluations

4 Model Evaluations

Automated Evaluations: To evaluate the grammaticality of the generated text, we used perplexity. We compared the text generated by the models against GPT Radford [2018] as the ground truth.

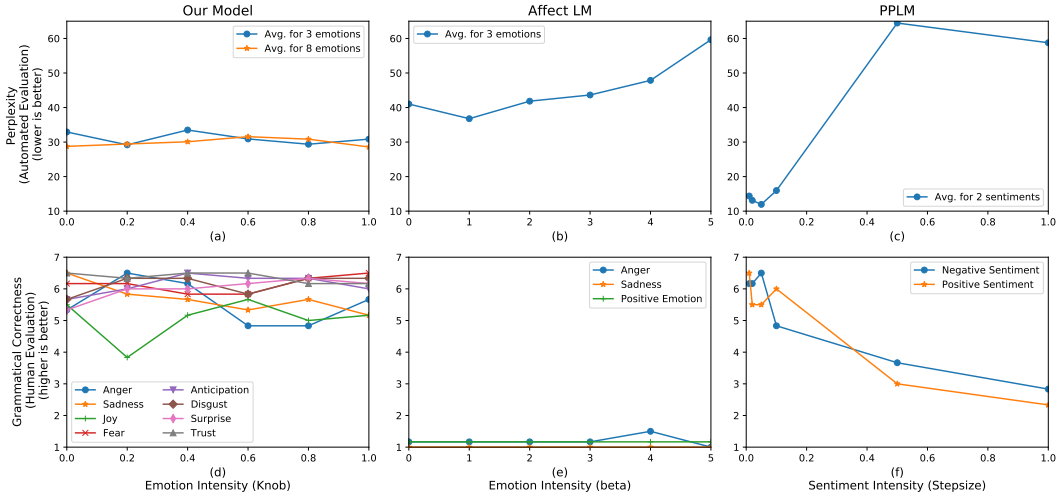


Figure 2: Grammatical Correctness Evaluations

Prompt	Topic	Emotion	Knob/ β	Text
Our Model				
The President	Politics	Anger	0.6	The President’s first budget is a huge tax cut for the rich and an increase in debt for our military. He’s a disaster. — Senator Bernie Sanders (@SenSanders) December 17, 2017..
The President	Politics	Anger	1.0	The President’s first budget is a huge tax cut for the rich and an increase in debt for our military. He’s a war criminal who wants to destroy our economy and take away our freedom.
AffectLM				
The President	NA	Anger	3	the president of the country you know . i i i don’t know i just hope that it ’s just . it ’s just going to kill innocent people .
The President	NA	Anger	5	the president of the country the fight against the united states hate the us and mass destruction attack and fight against killing you hate hate hate hate hate

Table 1: Examples of sentences generated by our model and AffectLM.

Human Evaluations: We conducted human evaluations comprising over 400 sentences. Two graduate students fluent in English (and oblivious to this project) evaluated sentences generated by our model and other competitor models (AffectLM, PPLM). Annotators were given sets of 6 sentences generated for an emotion class at different intensities. They were informed about the input emotion category of the set. They were asked to rate grammatical correctness for each sentence on a 7 point Likert scale, and to rank these sentence from 1 to 6 based on the relative intensity of the emotion expressed, 6 being the highest intensity rank. Averaged ratings from both the annotators are presented in Fig.1 and Fig.2. We observed great agreement ($\alpha = 0.74$) on grammatical correctness ratings, while we got $\alpha = 0.23$ on intensity rankings, showing the diversity in perceived emotion intensities. This is due to the subtle differences in intensities and diversity observed in the generated text.

5 Results and Conclusion

From Fig.1, we can see that for our model, the average perceived intensity across all emotions increases with model intensity, and then decreases slightly towards the end while that for AffectLM (Fig.2) doesn’t decrease at higher intensities. From Table 1, we see that our model intelligently models emotion and its intensity while optimizing for the grammar, while AffectLM just increases affect words without caring about grammar. Fig.2 proves the consistency of our model in generating grammatically valid sentences. We can conclude that our model is able to manipulate the emotional intensity of the generated text explicitly for certain emotions, and subtly for the remaining emotions (attributed to the subjective nature of emotion perception), while conveying the desired meaning.

References

- P. Ammanabrolu, W. Cheung, D. Tu, W. Broniec, and M. O. Riedl. Bringing stories alive: Generating interactive fiction worlds, 2020.
- Y. Bisk, A. Holtzman, J. Thomason, J. Andreas, Y. Bengio, J. Chai, M. Lapata, A. Lazaridou, J. May, A. Nisnevich, N. Pinto, and J. Turian. Experience grounds language, 2020.
- P. Colombo, W. Witon, A. Modi, J. Kennedy, and M. Kapadia. Affect-driven dialog generation. In *Proceedings of the 2019 Conference of the North*. Association for Computational Linguistics, 2019. doi: 10.18653/v1/n19-1374. URL <https://doi.org/10.18653/v1/n19-1374>.
- S. Dathathri, A. Madotto, J. Lan, J. Hung, E. Frank, P. Molino, J. Yosinski, and R. Liu. Plug and play language models: A simple approach to controlled text generation. In *International Conference on Learning Representations*, 2020. URL <https://openreview.net/forum?id=H1edEyBKDS>.
- S. Ghosh, M. Chollet, E. Laksana, L.-P. Morency, and S. Scherer. Affect-LM: A neural language model for customizable affective text generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 634–642, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1059. URL <https://www.aclweb.org/anthology/P17-1059>.
- J. Luketina, N. Nardelli, G. Farquhar, J. Foerster, J. Andreas, E. Grefenstette, S. Whiteson, and T. Rocktäschel. A survey of reinforcement learning informed by natural language, 2019.
- R. A. Mar, K. Oatley, M. Djikic, and J. Mullin. Emotion and narrative fiction: Interactive influences before, during, and after reading. *Cognition & Emotion*, 25(5):818–833, Aug. 2011. doi: 10.1080/02699931.2010.515151. URL <https://doi.org/10.1080/02699931.2010.515151>.
- S. M. Mohammad. Word affect intensities. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, Miyazaki, Japan, 2018.
- A. Radford. Improving language understanding by generative pre-training. 2018.
- A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. Language models are unsupervised multitask learners. 2019.
- T. R. Sumers, M. K. Ho, R. D. Hawkins, K. Narasimhan, and T. L. Griffiths. Learning rewards from linguistic feedback, 2020.

Appendix

A Examples Generated by our Model

Prompt	Topic	Emotion	Knob	Text
There exists	Science	Fear	0	There exists a vast, vast universe of material that is known as the material record, or MDR, that we can study and understand.
There exists	Science	Fear	0.4	There exists a theory of history that claims the United States, the most modern society in its history, was created by a race of people genetically altered in the laboratory.
There exists	Science	Fear	0.6	There exists a very good argument that the world is going to end, or is about to end. It's a frightening idea, and I've always felt that we are all doomed.
There exists	Science	Fear	1.0	There exists a very good argument that the US has been using chemical weapons against the population since the 1970s. In the case of chemical weapons, the evidence is quite overwhelming.
Once upon	Religion	Joy	0	Once upon a time, there were two types of people in America: those who were born into wealth, and those who were born into poverty.
Once upon	Religion	Joy	0.4	Once upon a time, a man named David had a dream. His dream was to be an inventor. The dream had a simple explanation. It said, "I want this."
To conclude	Politics	Anger	0	To conclude, the evidence is clear that the US has the ability to make its own decisions regarding what to do with the oil and gas in the ground, which is why we do this.
To conclude	Politics	Anger	0.2	To conclude, the evidence is clear that the US government is engaged in a systematic campaign of disinformation and propaganda against Russia.
I feel so	No topic	Sadness	0	I feel so sad for the girls that are still in high school. I'm sure that they will be able to go on to college without any problem.
I feel so	No topic	Sadness	0.4	I feel so bad for the kid that was killed. The guy was a great guy. He was the best of his age, he was just trying to do his thing.
Prior to this	Technology	Anticipation	0.6	Prior to this year, I had been a fan of the band. But this year, after spending a few days with them on the road, I'm really excited to get my hands on their new album 'Sleeping With the Enemy'.
Prior to this	Technology	Anticipation	1	Prior to this year, I had been a fan of the Nintendo Switch for over a year before it was released, but I've been waiting on the console to release and I'm excited to finally have it available to me.
To summarise	Politics	Disgust	0.6	To summarise: the current system of government is not designed for the economy of large, complex economies, but is designed to allow governments to do what they like to do best and that is to abuse their authority by abusing citizens.
To summarise	Politics	Disgust	1	To summarise: the current system of tax planning is flawed and needs reform. There is a growing body of evidence that the current taxation systems of many countries are deeply unfair and distortive of economic activity.
The pizza	No topic	Surprise	0	The pizza place is now open to the public! The pizza place is now closed and we are no longer accepting reservations.

The pizza	No topic	Surprise	0.4	The pizza place is now open to the public! The pizza place is now closed and we are no longer serving pizza! It is now open to the public! The pizza place is closed.
More importantly	Legal	Trust	0.6	More importantly, they're going to be getting a lot of free food. That will be the most important thing. You don't want to give them anything they don't need.
More importantly	Legal	Trust	1	More importantly, they're doing it in America, which is one of the greatest bastions of free speech in the world. The First Amendment protects freedom of assembly and expression.

B Examples Generated by AffectLM

Prompt	Topic	Emotion	Knob	Text
There Exists	NA	Anxious	3	there exists in the world . oh i bet . we think . oh yeah . oh you 're worried about it . yeah fear factor . yeah . yeah .
There Exists	NA	Anxious	5	there exists . you worry death factor death death factor factor fear factor factor fear factor fear factor fear factor
The President	NA	Anxious	0	the president of the country . yeah . i live in a small town . mhm . and we have to keep 'em . oh that 's great . and i don't know i can i can just imagine how you know if you know you 're just gonna have to have you know .
The President	NA	Anxious	2	the president of the country that 's in the in the middle east . oh yeah . and the terrorist attack is being in the united states . uhhuh . and ah uh .
Once upon	NA	Positive	0	once upon a time to get around the airport . i'd be willing to do that . okay . i would say that i went to the airport a lot and i was fortunate in the early seventies and we just went back and we went to israel and we did and there was a lot of people who were just to come over and you know they took it like there was something wrong with the government like the war that
Once upon	NA	Positive	2	once upon a time . i think you have to do it for two hours . yeah it 's it 's easy it 's good and it 's it 's wonderful . it is it is it 's very important too .
The Pizza	NA	Positive	0	the pizza . mhm . oh the pizza . oh okay . i don't know if you know where you 're coming from but yeah it 's not it 's not that familiar with the world .
The Pizza	NA	Positive	2	the pizza . yeah . yeah so i guess it 's a nice place to eat at home . yeah yeah i have a sister in law that 's out here .
To review	NA	Anger	0	to review the ah . the the religious israeli . mhm . ah the religious leaders of the country uh . and ah . ah we have ah um . uh .
To review	NA	Anger	2	to review the whole thing . yeah . well . you know we haven't had anything to do with it . at that point of september us we don't want to be in the war and you know i hate to say that we 're going to fight it you know .
In brief	NA	Anger	3	in brief like in a situation like that i think it would be a a a a a not a threat to the united states but i think it should be that way .

In brief	NA	Anger	5	in brief fight . i hate hate hate hate hate hate hate hate hate hate hate kill hate fighting kill mass mass destruction i hate hate hate hate hate hate hate hate hate hate hell .
I want to	NA	Sadness	0	i want to see what type of thing . you know i mean if if i don't want to learn how to speak i speak to my best friend . mhm . and if i can't speak with my brother i 'm not going to speak english right now .
I want to	NA	Sadness	2	i want to be my friends . so i was thinking about it for a few days they had they were having terrible feelings about that . yeah . yeah . you know so that 's the sad thing when you 're in that situation .
Views on	NA	Sadness	3	views on things . i don't know i think i think someone that 's losing their feelings from the poor end of the world you know .
Views on	NA	Sadness	5	views on losing feelings feelings in feelings feelings feelings about feelings feelings about gossip . gossip hurts hurt feelings