

This week I am looking for creepiness metrics.

---

My general idea looks like this. We take a story as input. Then it will generate emotional vectors and SBERT sentence embedding for every single sentences.

Then each of them will go through LSTM models. One of them has 26 neurons and two LSTM layers with drop out. Another has 768.

Finally, we concatenate them into one vector as out output.

---

But how do we know what emotions to choose from? After some googling, I found this paper by Alan and Dacher in 2017. Their results were discovered after gathered 2,185 short videos showing a wide variety of situations and events. The videos were chosen to provoke particular emotions from more than 800 participating viewers. They concluded that we essentially have 27 emotions.

In particular to our project, fear and horror are emotions we are interested in. Also, I later found out anxiety and empathetic pain are also somewhat related to creepy stories.

To find the emotions of a text, I am using cosine similarities.

---

Here are some examples sentences from creepypasta. One of the sentences said, "Pinned to the walls were decaying bodies of young men and women, dressed in all white outfits." It gives the text a really high horror score.

Here is another one, "I held my breathe as ran to the door". We can see that it matches the emotions pretty accurately.

---

Here are more examples. We can see that it not only detect horror, but also detect it as empathetic pain. I think empathetic pain is more accurate than horror because when a person held a knife and cut his fingers, maybe he is cooking.

Anyway, it is recognizing emotions somewhat accurately.

---

So I run through each sentence in a story and record their emotion scores.

---

Here is the horror level across the most upvoted story in r/confessions. This plot's y upper limit is 0.6. It doesn't have many sentences, but its max horror score is around 0.5.

Compared to the most upvoted story in r/nosleep, we can see that some of the horror scores are through the roof.

---

On the other hand, if we plot all the emotions in one plot. I made a gif here. I am sorry that x ticks were cut off. But the vertical blue lines are Anxiety, empathetic pain, fear and horror as I mentioned before. I plot 5 sentences each frame in this gif.

Compare these two stories, we can see that the creepy story sometimes can get really high along the blue line.

---

However, when I wanted to train the LSTM model, I ran into a problem where I don't have enough rams. LSTM model and our data take so many rams. I have 16 gb of ram and 8 gb of vrams yet it's still not enough.