

Joke Generation Using Masked Language Infilling on Automatically Extracted Templates

Mayank Goel Abhijit Manatkar
IIIT Hyderabad

mayank.goel@research.iiit.ac.in abhijit.manatkar@students.iiit.ac.in


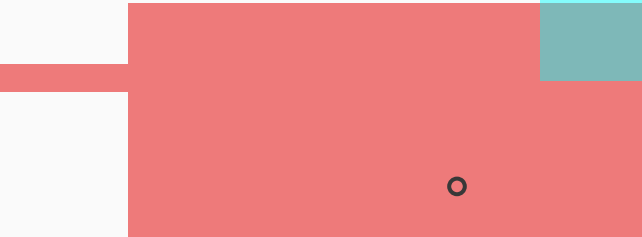


Introduction






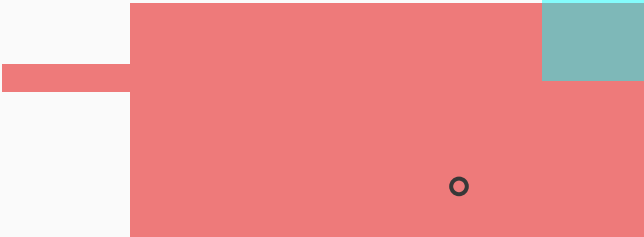



Background

- Humor is an important task to help us understand human cognition and language
 - While jokes are a normal part of society, AI is far from understanding jokes
 - We aim to generate jokes using a novel method, based around intuitive ideas of jokes
 - Utilizing the current state of art in NLP, we aim to understand the progress in understanding Language
- 
- 





Existing Literature

- Chaudhary, Tanishq, Mayank Goel, and Radhika Mamidi. "Towards conversational humor analysis and design." is the first attempt at extracting templates from jokes and then infilling them using BERT. Template extraction uses a heuristic formula based of human-tuned scores for POS tags and word frequency
- 
- 
- 
- 
- 



Existing Literature

- Pun generation from a pair of homophones [He et al., 2019] try to model the surprise which a word has in the context
- Joke generation using templates and rated jokes, depends on traditional ML methods [Winters, Nys, and De Schreye., 2019]
- LSTM Models for joke generation [Cai, Li, and Wan., 2018]






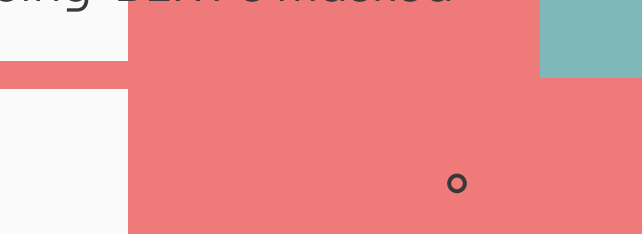
Understanding Humor

- Incongruity Theory of Humor where a joke arises out of the “incongruity” or disharmony of concepts, the resolution of which leads to jokes
- Two jokes can have similar structures (templates), but differ in the content that leads to a joke being incongruous
- Can we utilize this existing structure to generate novel content?





Structure of a Joke

- For the purposes of this presentation, we make the distinction between a joke's Style and Content.
 - Style is the template of a joke, it includes the stopwords and all other words that create the structure of the joke
 - Content are words with a high semantic contribution that fit into a joke to make it funny
 - This project looks at separating a joke into its style and contents, and replacing the contents with new words using BERT's masked language modelling
- 
- 





Methodology






Overview

- Take an existing joke, Identify content chunks using regular expressions on POS Tags
- Calculation semantic weightage of each chunk
- Mask tokens according to semantic weightage, based on controllable parameters
- Infill the masked tokens using fine tuned distillbert
- A new joke is generated





Content Chunk identification


- We identify chunks of words from the jokes which contribute to the “content” of the joke.
 - To identify these content chunks, we use a heuristic based on semantic similarity calculated using cosine scores from SentenceBert embeddings.
- 



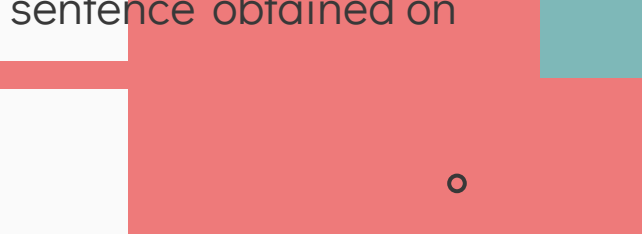



Content Chunk Identification

We apply regex to extract any chunk following any of the following patterns in the original joke:

- ADV* ADJ*
 - PDT? DET? ADJ* NOUN*/PROPN*
 - PDT? DET? ADV* VERB*
- 

Once these chunks have been extracted, we estimate the semantic contribution of each of these chunks by comparing SentenceBert embeddings of the original joke to SentenceBert embeddings of the “sentence” obtained on removing one of the chunks.





Template Extraction

- Once the chunks with highest semantic contribution have been identified, we replace them with masks. We experiment with two strategies for masking:
 - Mask top $n\%$ of content words from the sentence
 - Mask chunks contributing upto $p\%$ of the content
- We observe that setting higher values for n or p results in templates with more masks. For eg:
 - Original Joke: *Why didn't the sun go to college? Because it had a million degrees*
 - Template @ $n=0.3$: *Why didn't [MASK] go to [MASK] ? Because it had a million [MASK]*
 - Template @ $n=0.6$: *Why didn't [MASK] [MASK] to [MASK] ? Because it [MASK] [MASK] million [MASK]*



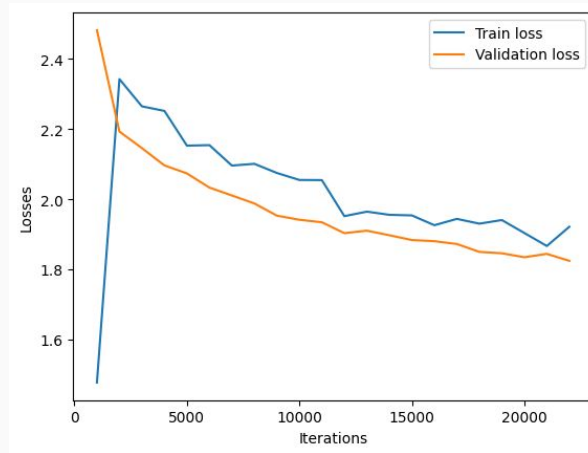
Fine Tuning BERT for Jokes - Dataset

- In order to get the most relevant in-filling, we fine-tune a distillbert model on a corpus of jokes consisting of jokes from Reddit r/jokes, the short jokes dataset and taivop's joke dataset
 - After removing duplicates and filtering for length, we remain with a corpus of 267,330 jokes, of which we use 80% for fine-tuning and keep aside 10% as a development set and a further 10% as the test set.
- 
- 




Fine Tuning BERT for Jokes

We finetune a distillbert-base-uncased model on the masked language modelling task using examples from the training set. We train for 2 epochs using a batch size of 16 and learning rate of $2e-5$.





Masked In-filling

- BERT is trained on Masked Language Modelling, which is the task of masking some of the words in a sentence and predicting which words should replace those masks
 - Thus, BERT is well-suited for generating new tokens given masked tokens.
- 





Example

Original Joke: What does a detective wear on a flight? Plainclothes

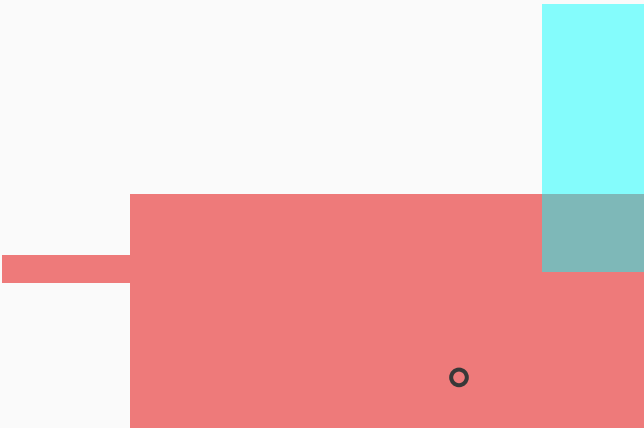




Example

Original Joke: What does a detective wear on a flight? Plainclothes

Chunks: a detective, detective, wear, a flight, flight, Plainclothes



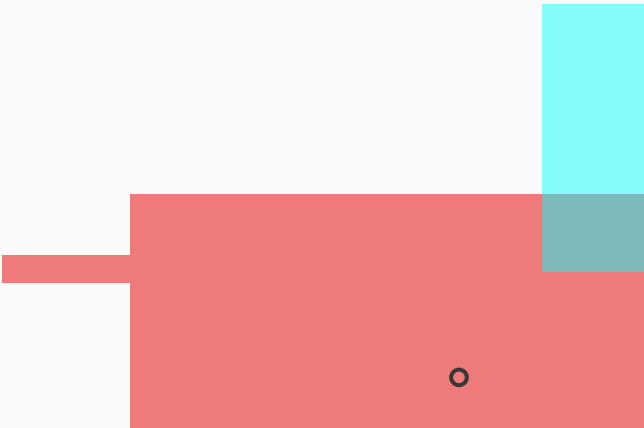


Example

Original Joke: What does a detective wear on a flight? Plainclothes

Chunks: a detective, detective, wear, a flight, flight, Plainclothes

Masked template: What does [MASK] wear on [MASK] ? Plainclothes





Example

Original Joke: What does a detective wear on a flight? Plainclothes

Chunks: a detective, detective, wear, a flight, flight, Plainclothes

Masked template: What does [MASK] wear on [MASK] ? Plainclothes

New joke: What does Santa wear on Friday? Plainclothes





Analysis



Results

Original	Masked	New	Comments
What kind of underwear does John Grisham use? Pelican Briefs I'll go find a bridge	What kind of [MASK] does [MASK] use? [MASK] I'll go find a [MASK]	What kind of perfume does Hitler use?? I'll go find a cure	Input data cannot be fully clean, can lead to faulty masking and structure
Why do sumo wrestlers shave their legs? To avoid being confused with feminists.	"Why do [MASK] [MASK] their [MASK] ? To avoid being confused with [MASK]	Why do women smoke their weed? To avoid being confused with alcohol.	Well-extracted templates allow for novel jokes

Results

Original	Masked	New	Comments
What did Buddha say to the hot dog vendor? Make me one with everything.	What did [MASK] [MASK] to [MASK] ? Make me one with everything.	What did Hitler propose to Stalin? Make me one with everything.	Model attempts to use words frequently seen in training data (Hitler etc)
What do you call a black man flying a plane? A pilot you racist	What do you call [MASK] [MASK] flying a plane? A pilot, you [MASK]	What do you call an Mexican flying a plane? A pilot, you racist	Similar intuitions can be maintained, while replacing the content

Human Evaluation


To study whether jokes generated by our method are able to pass for real jokes, we conducted a human evaluation in which we randomly presented either a human written joke or a generated joke to the evaluator and they are asked to identify whether the joke is human-written or machine generated.

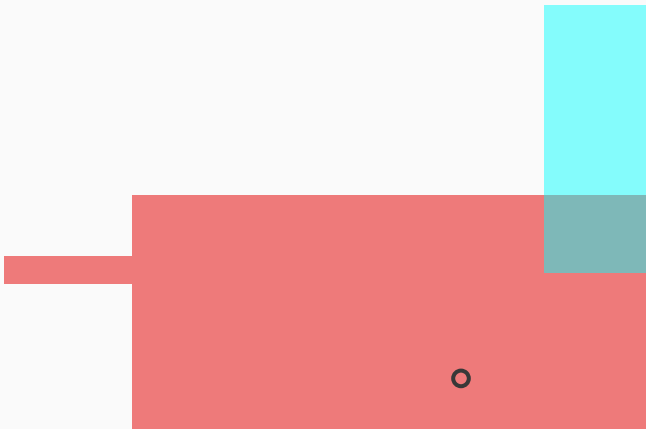
We performed this evaluation among 4 human evaluators and obtained the following results:

Joke Type	Reported as Real	Reported as Generated
Human	34 (68%)	16 (32%)
Generated	14 (24.56%)	43 (75.44%)




■ Limitations and Future Work

- Template Extraction is heavily dependent on heuristics. Better understanding of style/content separation in sentences can lead to better masking
 - Masked infilling is currently trying to increase semantic distance from the previous iteration, using something like a knowledge graph can lead to better jokes
- 





References

- Chaudhary, Tanishq, Mayank Goel, and Radhika Mamidi. "Towards conversational humor analysis and design." arXiv preprint arXiv:2103.00536 (2021). 
- Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
- Hossain, Nabil, et al. "Stimulating creativity with funlines: A case study of humor generation in headlines." arXiv preprint arXiv:2002.02031 (2020).



References

- Valitutti, Alessandro, et al. "“Let everything turn well in your wife”:
generation of adult humor using lexical constraints." Proceedings of
the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2013.
- Zhang, Hang, et al. "Let's be Humorous: Knowledge Enhanced Humor
Generation." arXiv preprint arXiv:2004.13317 (2020).
- Winters, Thomas, Vincent Nys, and Danny De Schreye. "Towards a
general framework for humor generation from rated examples."
http://computationalcreativity.net/iccc2019/assets/iccc_proceedings_2019.pdf Proceedings of the
10th International Conference on Computational Creativity (2019): 0



References

- Petrović, Saša, and David Matthews. "Unsupervised joke generation from big data." Proceedings of the 51st annual meeting of the association for computational linguistics (volume 2: Short papers). 2013.
- Weller, Orion, and Kevin Seppi. "The rjokes dataset: a large scale humor collection." Proceedings of The 12th language resources and evaluation conference. 2020.





Thank you.

