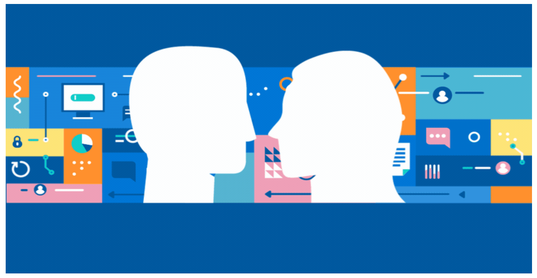
Generating Text through Adversarial Training using Skip-Thought Vectors

Afroz Ahamad
BITS Pilani, India

Introduction

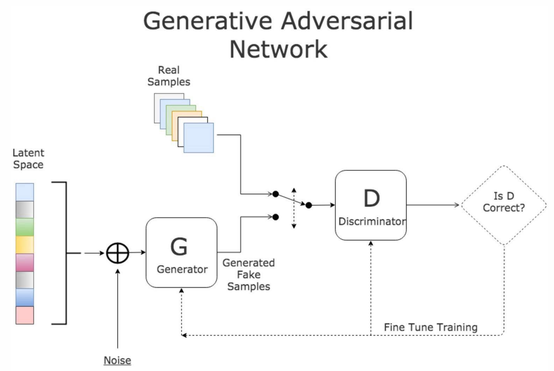
Inducing a style in generated text can lead to producing acceptable responses in:

- dialogue generation,
- image captioning and
- artificial chat bot systems.



- In literature corpora the vocabulary does not vary significantly across the authors, but the manner of expression does, which is intuitively best captured at the level of sentences.
- Make the adversarial model approximate the distribution of all sentences (rather than words or characters).

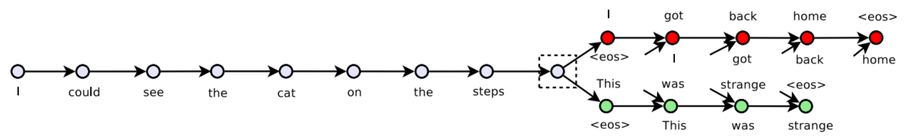
Motivation



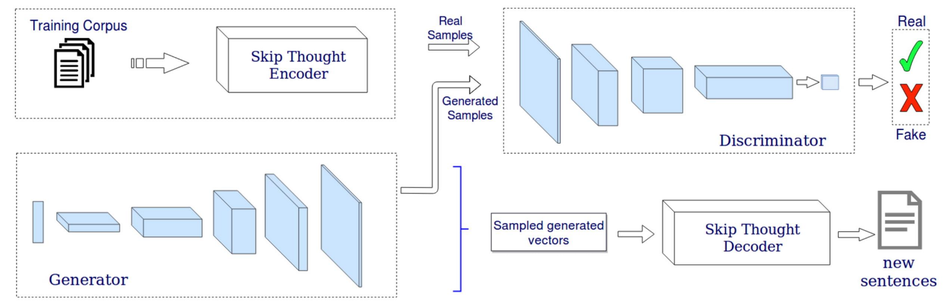
Attempts have been made to utilize GANs in different setups with word embeddings for text generation.

Proposed architecture aims to reproduce writing style in generated text by modelling the way of expression at a sentence level across all the works of an author.

Methodology



The skip-thoughts model.



Generator is trained with Skip-Thought sentence embeddings produced for the training corpus.

GRU decoding cells reconstruct sentences from vectors sampled from GAN generator network using greedy decoding.

Skip-Thought Generative Adversarial Network Architecture

For training of the STGAN, the Skip-Thought encoder produces sentence embedding vectors which are labelled as real samples for GAN discriminator.

The decoder uses greedy decoding by taking the argmax over the softmax output distribution for a given time-step which also acts as input for next time-step.

Skip Thought decoder reconstructs sentences conditioned on a sentence vector by randomly sampling from the predicted distributions with a preset beam width.

Conclusion and Further Work

- Presents a simple and effective model for text generation based on adversarial training using sentence embeddings.
- Discusses how the automated corpus-based evaluations correlate with human judgements.
- In future, this work aims to be applied for synthesizing images from text, exploring complementary architectures to projects like neural-storyteller.

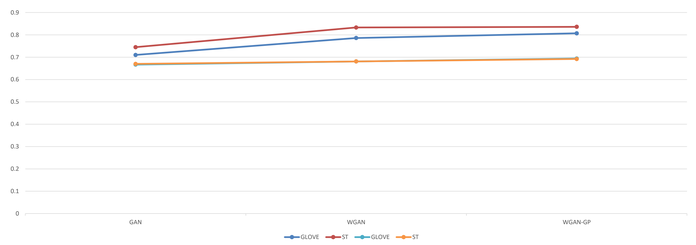
Results

Extensive experiments were run in different embedding settings on conditional text generation and language generation.

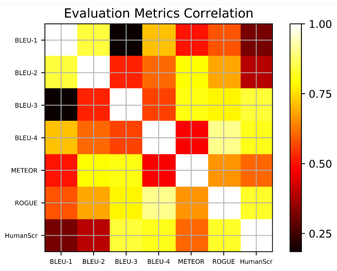
The model outperforms baseline text generation networks.

Model	Embedding	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE
LSTM	GloVe Average	0.874	0.792	0.621	0.582	0.681	**0.692**
	GloVe Extreme	0.874	0.791	0.616	0.580	0.677	0.685
	Skip Thought	**0.885**	**0.807**	**0.633**	**0.585**	**0.683**	**0.692**
Attention BiLSTM	GloVe Average	**0.904**	**0.836**	0.645	0.583	**0.695**	0.698
	GloVe Extreme	0.886	0.827	0.643	0.581	0.689	0.696
	Skip Thought	0.900	0.827	**0.651**	**0.589**	0.692	**0.715**
WGAN -GP	GloVe Average	0.879	0.807	0.668	0.585	**0.694**	0.702
	GloVe Extreme	0.853	0.799	0.666	0.579	0.689	0.697
	Skip Thought	**0.903**	**0.836**	**0.682**	**0.594**	0.692	**0.731**

Table 1: Evaluation of models on word-overlap based automated metrics when trained with different embeddings. Skip-Thought gives better results than GloVe for BLEU-n and ROUGE metrics, while the METEOR scores are comparable to that when using averaged GloVe embedding with Attention BiLSTM generator.



Skip Thought vectors vs GLoVe in different GAN setups.



Pearson's correlation coefficient between automated computed metrics and human scores.

Human scores correlate well with BLEU-3 and ROUGE scores.

Human Judgment Scores

Participants were asked to mark on a scale of 1 to 5 if they thought that a sentence seemed to belong to the author's works or was generated from a model,

	Real	Fake	% real	% fake
Real	30	**51**	37.04%	**62.96%**
Fake	**48**	75	**39.02%**	60.98%

Weighted human scores for sentences. |rating − 3| is weight given to each rating. 39.02% of the generated samples were marked as real.