A New Chatbot for Customer Service on Social Media

Presented By:
Rajat Kumar (201811024)
Nadeem Maulvi (201811048)

Guide: Surupendu G.

Supervisor:
Dr. Prasenjit Majumder
(Associate Professor)
DA-IICT

Authors of the paper [1]:

Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, Rama Akkiraju

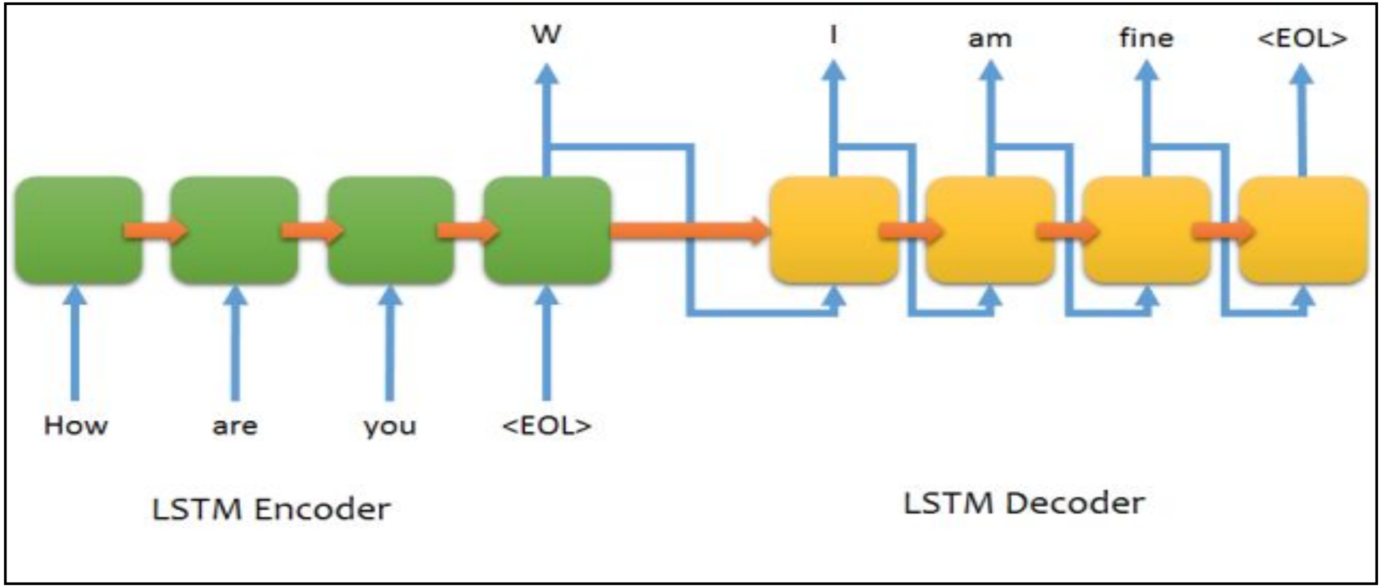
IBM Research - Almaden

San Jose, CA, USA

Abstract

1. The customers feel free and comfortable to post any thought or query on social media rather than email to a particular company. However, majority of these requests are not addressed timely or not even addressed at all.
2. Their system is trained on tweets taken from the twitter and is capable of showing sympathy to the customer as good as a human. The new conversational system is able to generate responses by itself. State-of-the-art deep learning techniques are used to implement the chatbot.
3. Dataset is taken from kaggle [2] which contains 0.1 million Question Pairs from Apple Support.

Model Used: LSTM layers



Model Specifications:

1. Number of LSTM layers used: 4 (Both for Encoder and Decoder part)
2. Hidden Layer Size: 640 (Word Embedding Size)
3. Number of Iterations: 70,000
4. Average Loss: 0.15 from 9.144
5. Back Propagation using Gradient Descent.

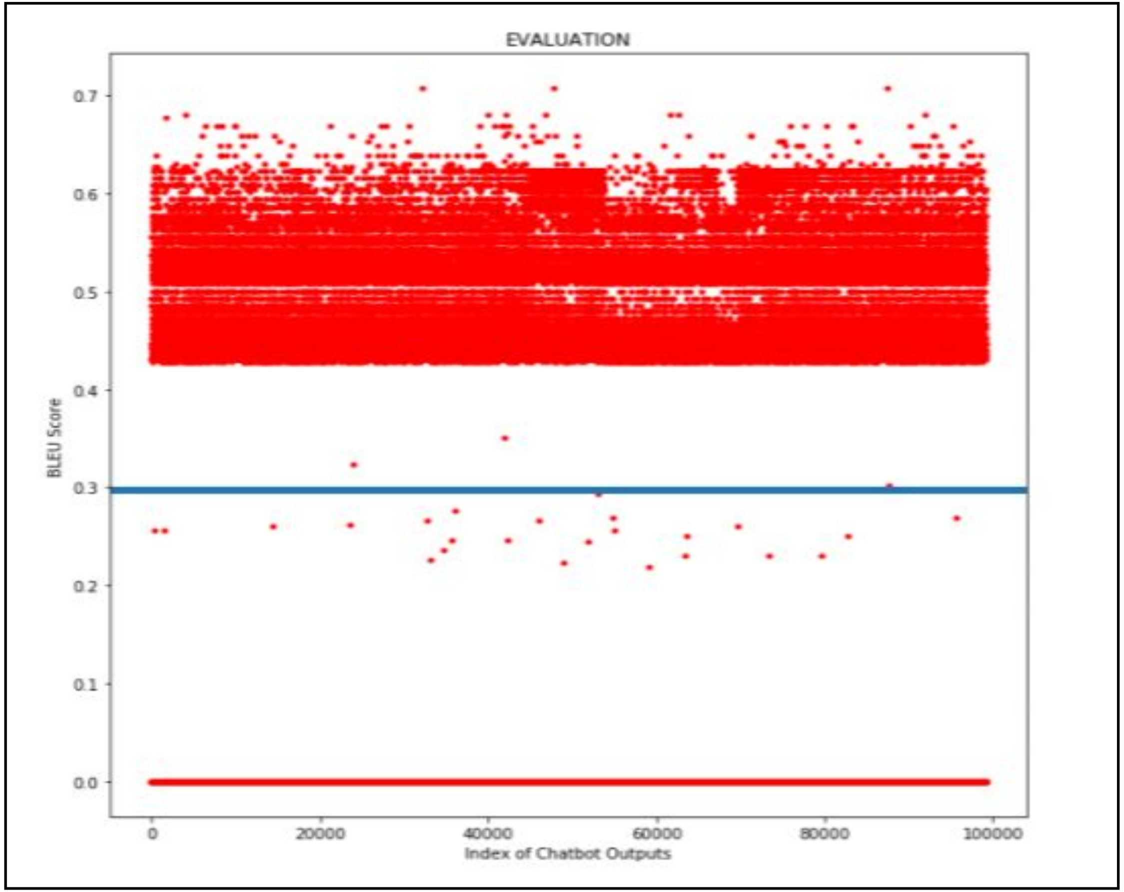
Automatic Evaluation- BLEU Score

Paper	[1]	**Our Score (Considering 0 BLEU scores)**	**Our Score (Not Considering 0 BLEU scores)**
BLEU Score	**0.36**	0.3	**0.5**

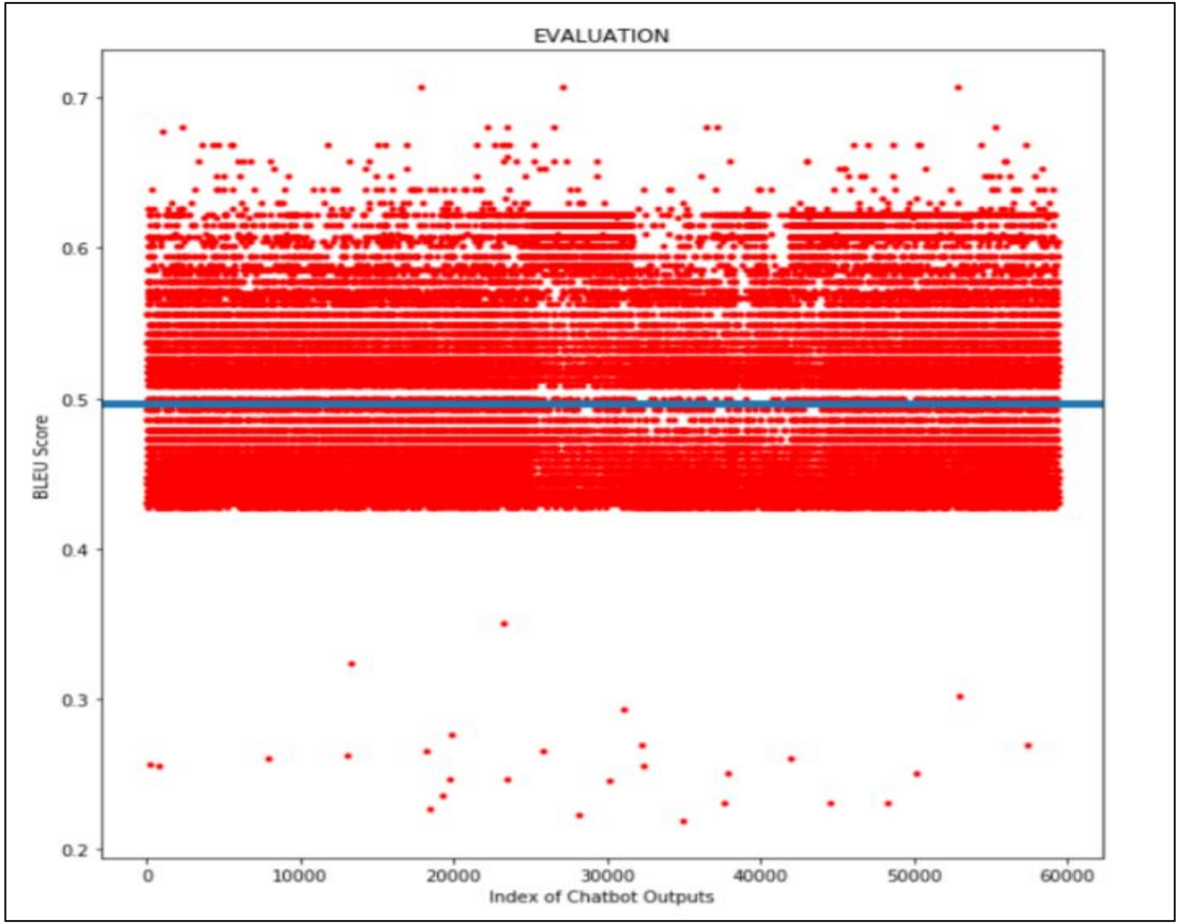
Human Evaluation (Using Google form results)

1. Google Form Contained 15 Chatbot Responses vs Actual Responses.
2. 22 User responses were recorded on a scale of 1 to 5.
3. The average Chatbot Human Evaluation comes out to be **3.578/5**

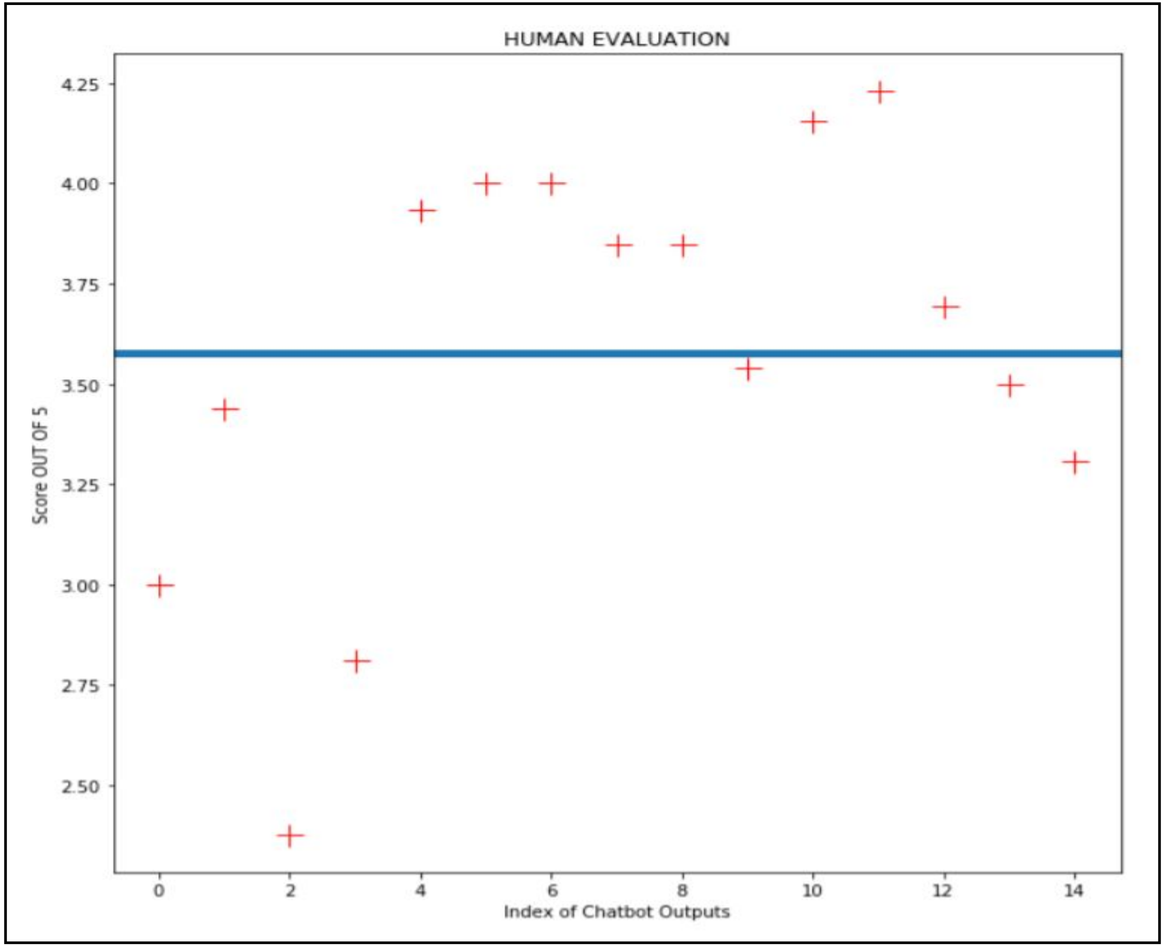
Visualizing Results- BLEU Score (Considering 0 Values) =0.3



Visualizing Results- BLEU Score (Without Considering 0 Values) = 0.5



Visualizing Results- Human Evaluation Score 3.578/5



Challenges

1. Computation Time is more so hyperparameter tuning can be done only after results are computed.
2. Global minima is not guaranteed always by Gradient Descent.
3. Human Evaluation expects results to be large scale.
4. BLEU score 0 values change overall BLEU score.

Conclusion and Future Work

1. After working on the project for more than 3 months, we gather insights about Encoder Decoder model.
2. Future work can be hyperparameter tuning for better results.
3. Better algorithms like simulated annealing can be used instead of gradient descent.
4. Better Evaluation metric can be considered.

References:

[1] Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. 2017. A New Chatbot for Customer Service on Social Media. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, USA, 3506-3510. DOI: https://doi.org/10.1145/3025453.3025496

[2] Dataset Link: https://www.kaggle.com/thoughtvector/customer-support-on-twitter

[3] Chatbot Tutorial Link: https://pytorch.org/tutorials/beginner/chatbot_tutorial.html

[4] Youtube Tutorial:
https://www.youtube.com/watch?v=CNuI8OWsppg