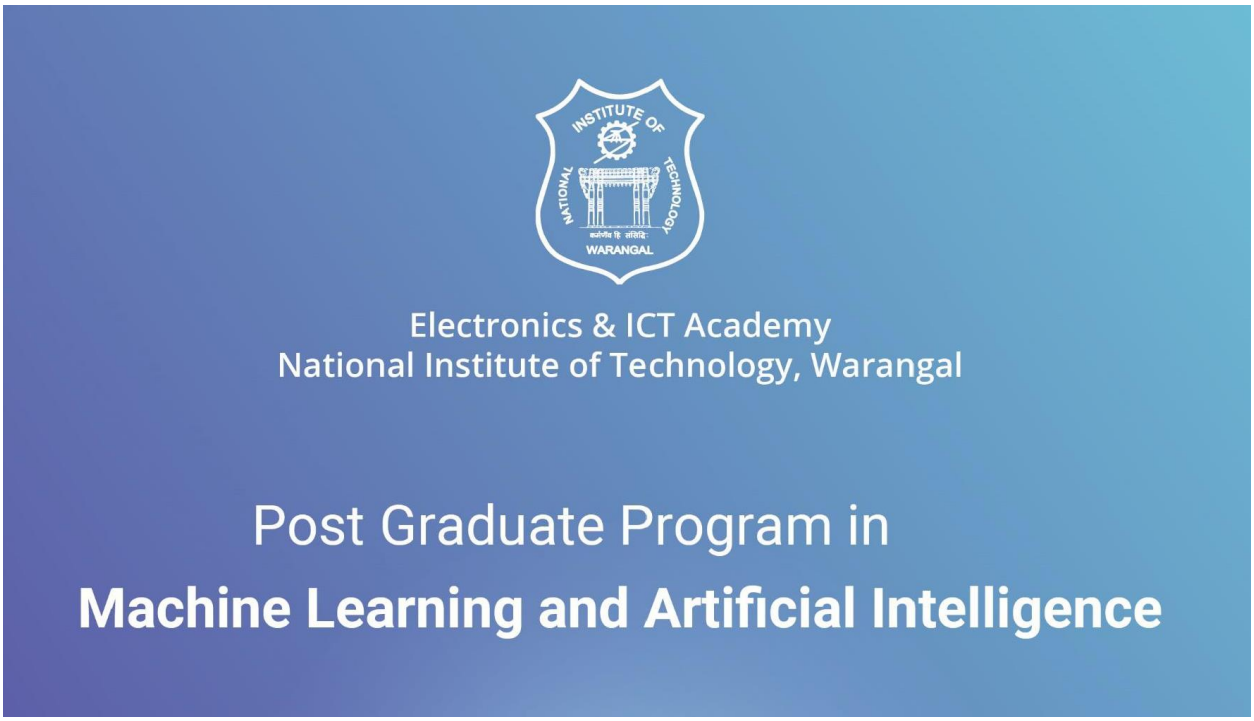


Project Synopsis

Submitted by

Jiss Peter

Technical Architect – Tata America International Ltd (TCS Ltd.)

Jiss07@gmail.com

+1(754)-248-3941

Abstract

This paper provides a probable solution for the chat bot creation problem statement as provided by Edureka and NIT Warangal by utilizing the LSTM encoder - decoder Architecture. I have used the architecture to show how this architecture is successfully applied to the challenging problem of defining a chatbot by processing the movie lines data as input.

1. Aim of the Project

Aim of the project is to build an intelligent conversational chatbot, Riki, that can understand complex queries from the user and intelligently respond.

2. Background

R-Intelligence Inc., an AI startup, has partnered with an online chat and discussion website bluedit.io. R-Intelligence Inc. has an average of over 5 million active customers across the globe and more than 100,000 active chat rooms.

Due to the increase in traffic, R-Intelligence Inc. are looking to improve the user experience by using a chatbot moderator.

The chatbot moderator uses a chatbot named Riki, that can help R-Intelligence Inc. customers to engage in meaningful conversation and keep them updated on trending topics. The Artificial Intelligence-powered chat experience provides easy access to information and a host of options to the customers

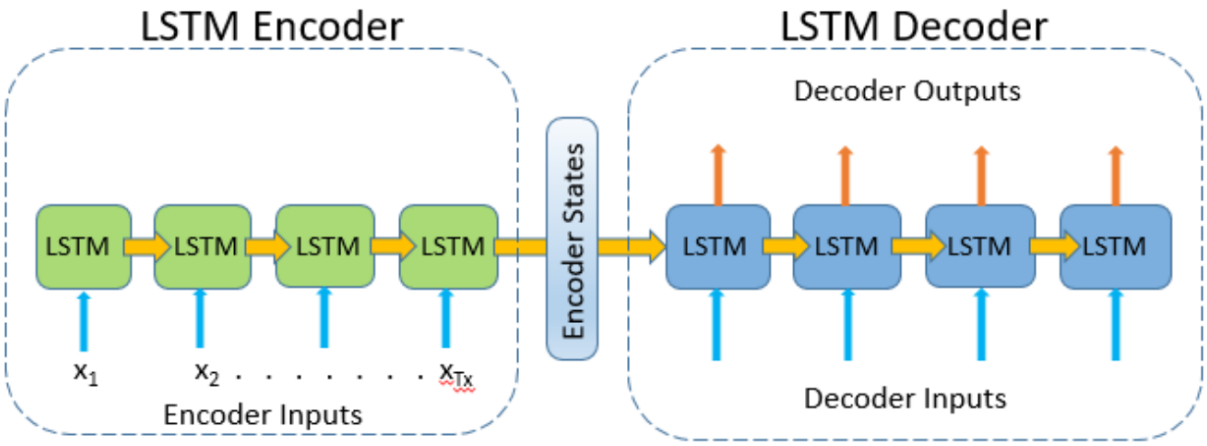
3. Problem Statement

R-Intelligence Inc. has invested in Python, PySpark, and Tensorflow. Using emerging technologies of Artificial Intelligence, Machine Learning, and Natural Language Processing, Riki – the chatbot should make the whole conversation as realistic as talking to an actual human.

The chatbot should understand that users have different intents and make it extremely simple to work around these by presenting the users with options and recommendations that best suit their needs.

4. Architecture Used and Solution Implemented

Long Short-Term Memory (LSTM) is a special kind of RNN, which has special forget gates, in addition to input and output gates of the simple RNN. The solution implemented for the given problem statement in the project is LSTM encoder-decoder Architecture. LSTMs are designed to remember the input state for a longer time than an RNN, hence allowing long sequences to be processed accurately. LSTMs are a fundamental part of NLP architecture for Apple, Amazon, Google and other tech companies.



The model consists of 3 parts: encoder, intermediate (encoder) vector and decoder. This is already being explained along with Model architecture.

The use of the models in concert gives the architecture its name of Encoder-Decoder LSTM designed specifically for seq2seq problems.

Sequence to sequence models are based on RNN architecture and consists of two RNNs: an encoder and a decoder. The encoder's task is to process the input, and the decoder to process the output. Sequence to sequence models can be thought of as one decoder node producing output corresponding to one encoder node.

This model has straightforward application in machine translation as a corresponding word for the output language can be generated by decoder easily by looking only at one word of input language at a time.

5. Real-World Applications

There are many ways to upgrade communication between your company and its customers. One effective method (both in terms of cost and results) for any business to improve their customer service game is by using chatbots. Recently, chatbots have been applied in many different aspects of business and have had many proven records of success

- Chatbots Answer Questions And Inquiries

A chatbot, thanks to its 24/7 presence and ability to reply instantly, can be of immense help. The company organizing the event can set up the chatbot to answer common questions like: What does the ticket cost? And where do I buy it? How do I get to the venue?

- Book Tickets To Events/Shows With Chatbots

Before making a purchasing decision, most customers will ask the same types of questions regarding what they are buying. Answering such repetitive questions will take up your customer support's valuable time and resources. A better solution would be to deploy a chatbot on your website and design it to answer basic questions your salespeople get regularly. These can be questions like: Which movies are playing today? Can I get a refund for my ticket? Which seats are available?

- Chatbots To Build Remarkable Customer Experience

Chatbots don't have to be serious and purely transactional. You can design them to go through a buying decision by creating a 'quiz', telling jokes along the way, and sending the occasional meme.

- Chatbots Can Confirm Orders And Track Shipping

After buying a product, customers may want to know how much time remains until their product ships. Chatbots Can be used to Confirm Orders And Track Shipping 24*7

Other applications includes

- Chatbots Help You Collect Customer Feedback Efficiently
- Chatbots Assign Customer Requests To Support Teams
- Chatbots Generate Leads With Sales Approach
- Chatbots Help You Build Email Lists More Effectively
- Chatbots Promote Products With Fun Conversations

6. Conclusion

We built an open domain generative chatbot using LSTM and seq2seq model with Tensorflow. We could be able to attain nearly same amount of value loss and accuracy loss during training which has resulted in to nearly 100% accuracy. Which shows that the training and testing was just right, and is not over or under fitting. The bot is able to capture simple entity, but most responses are pretty generic. End to end model using LSTM architecture is easy to implement, but using it alone can't yield good result. In the future we plan to try entity recognition and larger dataset with distributed training, and might switch to retrieval based chatbot. Also the open source frameworks such as RASA or RASA-X or Google's Dialogue-Flow provides better chat bot development environment as these frameworks provide complete built in architecture defined properly with Entity recognition with limited amount of coding.

7. Summary

Deep Learning is a new, exciting domain with tremendous research being carried out in this space. This has allowed us to create generic models to analyze large data, at scale. Deep Learning helps eliminate the complexity of feature engineering from the traditional machine learning process and truly learns the underlying data patterns. Architectures like RNN, LSTM and sequence to sequence model overcome the limitation of context recognition, an essential part of NLP. Generation-based Networks have it made possible to create a "true" chatbot whose responses are based only on the training data. These advances can be leveraged and explored to build life-like chatbots which can make real conversations with a human. By creating robust chatbots which can be customized according to the training data, large scale automation is possible.

8. Technology and Tools used

- Python 3 in Google colab environment with a 24 GB of GPU with an available 8GB RAM
- Tensorflow, Keras, Model and classification modules as available in Python

9. References

[1] Vyas Ajay Bhagwat - Deep Learning for Chatbots - San Jose State University

[2] Snowbot: An empirical study of building chatbot using seq2seq model with different machine learning framework by Pinglei Guo, Yusi Xiang, Yunzheng Zhang, Weiting Zhan.

[3] SnehaKudugunt - Deep Neural Networks for Bot Detection- Indian Institute Of Technology Hyderabad

[4] Navin Kumar Manaswi -Deep Learning with Applications Using Python by Apress.com

[5] Online medias such as kaggle, medium, analytics vidya etc.