**Designing Conversational Agents for the NUST Website**

Namibia University of Science and Technology

Faculty of Computing and Informatics

Lecture: Dr. Irja Shaanika

Supervisor: Prof. Jose Quenum

1st Semester 2020

Ndapopile Fideria

Honours Degree course: Research Methodology (RIT812S)

Semester: 7

Student number: 217094139

[fideriajt@gmail.com](mailto:fideriajt@gmail.com)

Abstract

Contents

1. Introduction …………………………………………………………………..#

2. Related Work…………………………………………………………………#

**Introduction**

A conversational agent or well known as a chatbot is a software that conducts conversations with humans through text or text to speech. Many companies in the past recent years have released chatbots on their products to provide a better experience to their users such as Apple’ Siri, Google’s personal assistant and Microsoft’s Cortana. These newly improved chatbots, stem from the mother of all chatbots ELIZA [Weizenbaum, 1966]. Weizenbaum’s aim was to demonstrate that the communication between man and machine was superficial. His interest derived from the work done by Alan Turing in the 1950’s that according to Turing a machine could impersonate a human, be capable of thinking and be intelligent. Since then with a lot of new technology advancements an intelligent chatbot is ever so in reach.

Businesses such as restaurants and movies theaters have implemented domain specify chatbots for tasks such as making reservations and booking seats respectively. This traditional oriented dialog conversational agents are limited and cannot imitate human dialogs in all its facets, but otherwise do the intended job satisfactorily to some extent. Other chatbots are built to not only answer questions specific to one domain, but to hold a conversation about almost anything, which is a quite ambitious task that many have attempted. Artificial Intelligence has brought this goal closer to reality.

Approaches to develop chatbots have evolved and improved from using rules and templates that are hard coded. Artificial Intelligence (AI) has introduced models to combat the limitations and enhance the capabilities of previous chatbots architectures. A chatbot for the Namibia University of Science and Technology (NUST) would require AI models that are able to engage a student or parent that in both the institution’s information but as handle out of domain questions. To make the experience of the students’ guardians even grander, the NUST chatbot will be interacted with in the native dominant language Oshiwambo.

In this work, I make a step toward more consistent and relevant data-driven conversational agent for the institution NUST by proposing a transformer model architecture. Transformers have curbed the disadvantages with traditional seq-2-seq, Recurrent Neutral Networks and Long Short Term Memory Network Models.

**2. Related Work**

2.1 Recurrent Neutral for NLP

Rumelhart (1986) cited a paper that applied back-propagation algorithm to multi-layer neural networks, with his co-authors Hinton and Williams. With many experiments, they aimed to describe a new learning procedure for networks of neurons-like units like those of human cognition. This gave birth to the sematic cognition as well as learned domains which are still the fundamental of newer improved models of this age.

Recurrent Neutral Networks (RNN)

RNN’s implementation suffers a slight flaw. The model cannot handle long sequences very well. It is only limited to short sequences because the RNN gradient becomes smaller and smaller and eventually vanishes, that no learning is done [Hochreiter, 1991]. RNN find it hard to remember information seen multiple steps ago when the steps of unrolling increases.

Even with RNN’ faults, its models such as Sequence Vector Models have been used to develop chatbots that are able detected whether the human their conversing with is happy or sad, the sentiment in the text or speech they provide. This ability would be a great attribute of our chatbot.

Sequence-Sequence Model is an RNN based which receive input as sequence and output another sequence, they can be used for language translation. Being able to tell a human’s mood and have translation capabilities are great ideal attributes of a conversational agent, but there are other AI models that offer this attributes and curb RNN’s disadvantages.

2.2 Long Short Term Memory (LSTM) Networks

LSTM were developed as a solution to RNN’s short comings. They have a branch that allows past information to skip a lot of the processing of the current cell and move on to the next which allows the memory to be retained for a longer sequences.

Diagram

Introduces in 1997 by Sepp Hochreiter and Jürgen Schmidhuber, --🡪 cite

Replaced neurons, replaced

LSTM has a branch that allows past information to skip a lot of the processing of the current cell and move on to the next which allows the memory to be retained for a longer sequences (solved the RNN problem)

Disadvantage:

- LSTM are way slower than RNN to train, because input data needs to be passed sequencentially or serially, one after the other

and we need input of the previous state to make any operations on the current state (Dependecy), such sequential flow does not make use

of todays GPS (designed for parallel computation)

2. 3 Transformers Neural Architecture

Introduced 2017

The network employs an Encoder Decoder architecture much like Recurrent Neural Networks

The DIFF is that the input Sequence can be passed in parallel

eg translation eng to french

Encoder Components:

Attention - > what part of the input should I focus on? (USERS INPUT IN THE CHATBOT)

Feed Forwards Nets

//getting the notion of context

Decoder Components:

Self Attention Block - generates attention vectots for every word in the (eg french word, for ME IDK :(, the answers i guess, lawwwwd help me)

No disadvantages?

**3. Datasets in Use**

3.1 Dataset Scraping

There are plenty of pre collected datasets that could be obtained, but none solely on NUST. It takes a long time to obtain sufficient dataset for a chatbot’s intelligence, sufficient being a total of 500 thousand records at least. Most datasets are build based upon for example Reddit answers of a score greater than 2, which means that a bot build upon such a dataset will be able to answer almost anything random question, and will some will have a lot of opinions some possibly incorrect and outdated answers.

This works’ chatbot aim is to have it answer any question on NUST. To combat the curve of having to collect this data, and not have outdated data is to collect the data right off the official NUST website. Julia a high-level, high-performance, dynamic programming open source language provides packages to will scrap the NUST website with ease.

3.2 Preprocessing

Our dataset is collected through web scraping and thus is not in the format that which a model can be trained with. It has a bunch of html related code, it has links to other websites, images and other multimedia.

Preprocessing an essential step in data mining where data is transformed and prepared for mining procedure [Alasadi et al., 2017]. It goes the same for AI models, we want clean and prepared data for the models. For a chatbot the data will be structured in a questions and answers structure.

3.1 Oshiwambo English Dataset

There are plenty of datasets for international languages to English, but not many of native African languages. In fact there no Oshiwambo to English dataset. The parallel dataset that we intent to use we simply must collect.

The tidies work of creating the dataset is possibly made easier to some degree by a recently developed local Namibian application found on Google Play Store. The app is digital dictionary where users can search for what English certain words are in Oshiwambo and vice versa, as well as giving an exemplary sentence of how it’s used. Without entirely depending but using the app as a point of reference to compare to what actual Oshiwambo speaking people translate certain words to in English.

The dataset is based on basic communication speech, as well as an Oshiwambo domain of institutional educational words users might express. The chatbot will have to translate what the user says to English, find and answer in the data collected and scraped from the NUST website, then translate the answer to Oshiwambo.

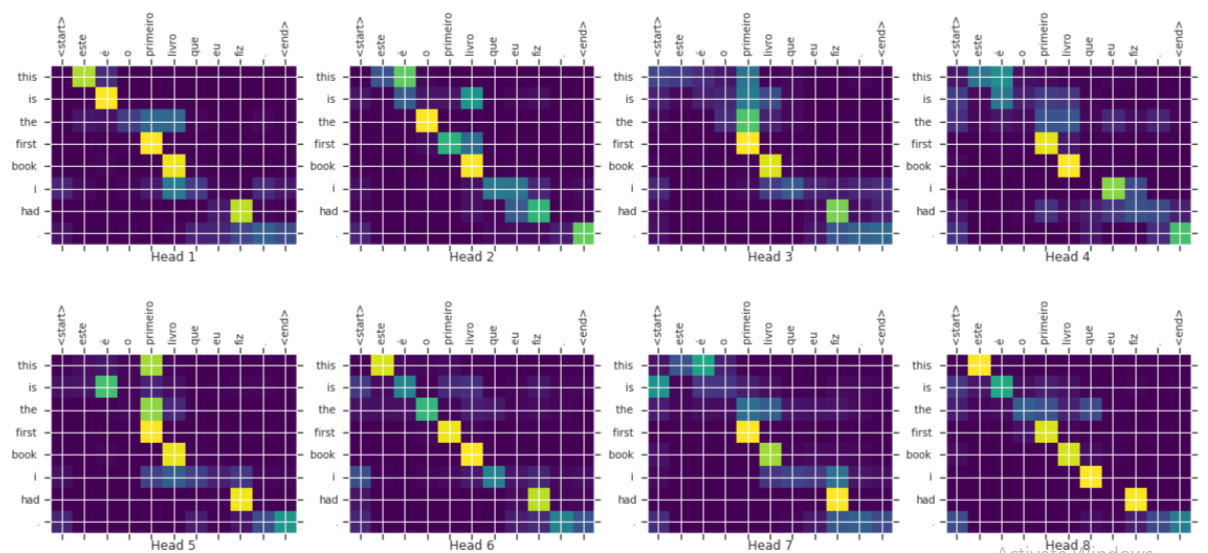
**4. Implementation and Evaluation**

To build a conversational agent for NUST, we will use Transfomers models. For a prototype of this chatbot a computer with aleast 6GB of processing memory and GPU, but with more of this resources the models work more efficiently.

4.1 Transformer model for language understanding

Tensoflow an open source platform that brings together developers and researchers to push and improve machine learning has lightly touch on a how a transformer model can translate portugues to English. They used Portugese-English translation dataset.

A transformer modle that translates portugues to English has been briefly touched on by Tensorflow. Transformers



**5 Experimental Setup**

**6 Conclusion**

6.1 Future Work

--Have I done any work yet?

**References**

[Apple, 2017] Apple (2020). Siri. https://www.apple.com/ios/siri/. Accessed:2020-15-22.

Weizenbaum, Joseph (1966). "ELIZA—a computer program for the study of natural language communication between man and machine". *Communications of the ACM*. **9**: 36–45. [doi](https://en.wikipedia.org/wiki/Doi_(identifier)):[10.1145/365153.365168](https://doi.org/10.1145%2F365153.365168).

Hochreiter, S. (1998). "The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions". *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. **06** (2): 107–116. [doi](https://en.wikipedia.org/wiki/Doi_(identifier)):[10.1142/S0218488598000094](https://doi.org/10.1142%2FS0218488598000094). [ISSN](https://en.wikipedia.org/wiki/ISSN_(identifier)) [0218-4885](https://www.worldcat.org/issn/0218-4885).

Alasadi, Suad A and Bhaya, Wesam S (2017). ["Review of data preprocessing techniques in data mining"](https://www.researchgate.net/profile/Suad_Alasadi/publication/320161439_Review_of_Data_Preprocessing_Techniques_in_Data_Mining/links/59d143d64585150177f3d15b/Review-of-Data-Preprocessing-Techniques-in-Data-Mining.pdf) (PDF). *Journal of Engineering and Applied Sciences*. **12** (16): 4102–4107.

Ayanouz, Soufyane & Anouar Abdelhakim, Boudhir & Benhmed, Mohammed. (2020). A Smart Chatbot Architecture based NLP and Machine learning for health care assistance. 10.1145/3386723.3387897.