

Cloud Based Machine Learning Chatbot

1st Sudheer Kumar Chakka
dept of Computer Science
Central Michigan University
Mount Pleasant, USA
chakk2s@cmich.edu or ORCID

2nd Shashank Kumar Reddy
dept of Computer Science
Central Michigan University
Mount Pleasant, USA
chint3s@cmich.edu or ORCID

3rd Amulya Gunturu
dept of Computer Science
Central Michigan University
Mount Pleasant, USA
guntu1a@cmich.edu or ORCID

Abstract—A Chat bot is a software application that is used to have a online conversation via text or voice. Chat can be with a automated trained machine or with an actual person [1]. There are many ways that a chat bot can be prepared and one of them is through machine learning. Using machine learning to create a chatbot that can process natural language is a challenging concept that includes to have a knowledge in Python, NLTK(Natural Language Tool Kit), NLP (Natural Language Processing), Machine Learning, Tensor Flow, Tokenize, Stemming, Bag of words, Random Shuffling, Google Cloud, Micro-services, and putty. A chat bot can be deployed in website or in cloud. In this paper will answer the following questions what is a chat bot? what are the types of chat bot? How to create create a chat bot? what do we need to create a chat bot? and finally How to deploy in Google Cloud? Goal of the project is to create a chat bot in python language with machine learning techniques and deploy in google cloud using micro-services [10].

Index Terms—Python, NLTK(Natural Language Tool Kit), NLP (Natural Language Processing), Machine Learning, Tensor Flow, tlearn, Google Cloud, Tokenize, lancaster Stemmer, Stemming, Bag of words, Random Shuffling, Micro-services and Putty.

I. INTRODUCTION

Alan Turing's article "Computing Machinery and Intelligence" proposed Turing Test as a criterion of intelligence. A computer program's ability to impersonate a human in a written conversation is the criterion for Turing Test. Joseph Weizenbaum's program ELIZA which is published in 1996 was able to fool users into believing them that they are conversing with a real human [1]. As per a report chat bots current value is 17.77 billion in 2020 and have forecasted an increase to reach 102.29 billion by 2026. Natural Language Processing(NLP) comes under Linguistic, Artificial Intelligence and Computer Science. NLP has its roots in the early 1950's i.e during the Alan Turing's article. The Georgetown Experiment in 1954 can be mentioned as a beginning of NLP. More than 60 Russian Sentences are changed into English automatically in this experiment [3]. Machine Learning is an subsection of Artificial Intelligence (AI). In the domains of artificial intelligence and computer gaming, an IBM employee Arthur Samuel is a pioneer. The study of computer algorithms that can learn and evolve on their own given experience and data is known as machine learning (ML) [3]. Google Brain's second-generation system is called Tensor-Flow. On February 11, 2017, version 1.0.0 was released. Tensor-Flow may run on several CPUs and GPUs, unlike the reference implementation,

which runs on a single device. Tensor-Flow is a machine learning and artificial intelligence software library that is free and open-source. It can be used for a variety of applications, but it focuses on deep neural network training and inference [4]. Google have announced an App engine, a platform that will be helpful for hosting websites and to develop them. This was announced on April 2008 but was became available generally on November 2011 [4]. Tokenization is one of the techniques used in machine learning that is eventually helpful in making the machine learn from the words that are split from sentences. Stemming is introduced and studied in computer science since 1960s. This technique is to get the root word for a given word so that we can analyse the actual meaning. Bag-of-words model is used in NLP and Information retrieval. Early reference for this can be found in 1954 from the article of Zellig Harris on Distributional Structure [5]. Micro-web-services are introduced by Peter Rodgers. Main work of this micro-services is to make sure that no two services are tightly coupled. Microservices is a form of application development that uses microservices architecture. Microservices allow a large application to be broken down into smaller, self-contained components, each with its own set of responsibilities [5]. Putty is a serial console, open-source terminal emulator that is free. This is a network file transfer application. Many software and networking companies use this to ease the login to servers and perform required work. This support Telnet, SSH, SCP, raw socket connection and serial port as well [6]. All the above mentioned are used in this project for implementation and are explained deeply in upcoming sections.

II. TYPES OF CHATBOTS

A. Machine learning Chatbot

Machine learning chatbot is a computer mimicking a human and speaking with human. A computer can mimic a human after it is trained with NLP, tokenizing, stemming, lematization, bag of words techniques. There are many types of learning inside machine learning like Supervised learning, Un-supervised learning, semi-supervised learning, Reinforcement Learning, and Dimensionality Reduction.

Unsupervised learning methods take a collection of data with only inputs and detect structure in it, such as data point grouping or clustering. The algorithm will learn from test data which is not classified, labelled or categorized. This identifies

commonalities in data and react. This algorithm created cluster analysis that is to create a subset of data based on similarity to one or more pre-designated methods.

We have used Supervised Learning for the chatbot. supervised learning algorithm creates a mathematical model from data which has both inputs and outputs which is known as Training data. This data will have an example of input and respective output which is also called as Supervisory Signal. The training example is represented as an array or vector called feature vector in the mathematical model. A matrix is used to represent training data. Supervised machine learning uses iterative optimization of an objective function to create a prediction based on provided training data and predict a possible output for a new input.

B. TF-IDF Chatbot

Term frequency and Inverse document frequency is referred as TF-IDF chatbot. Term Frequency is a simple count of counting the specific word occurrence in a given document. Inverse document frequency is checking the rarity of words. We count in how many documents a word is appeared. Like if a word is present 10 times in a document and in 100 documents, we count that the word is occurred only 100 times not 100×10 times. This converts the words into vectors and have a count of them.

C. Cosine Similarity Chatbot

A concept of vector space can be efficiently used in cosine similarity. We use the concept cosine similarity because when we have different length of documents then the result might be wrong. So we have to use length normalization concept and here comes the Cosine Similarity. This is a monotonically decreasing function ranging from 1 to -1.

The general concept for building the chat bot would be assume each sentence as a document on its own, when a user inputs a chat, consider it as a query. Then use the concept TF-IDF and cosine similarity. Have a comparison between two concepts and get most similar sentence from data set and provide it as output.

III. TECHNOLOGIES AND TECHNIQUES

Building a chat bot needs knowledge in many Techniques and deploying in cloud needs knowledge in Google Cloud and Micro-services.

A. Python

Python is a high level object oriented programming language, it is dynamically typed and garbage collected. Python supports programming paradigms, functional programming and object oriented. Because of comprehensive library python is often described as batteries included. Python is very easy to use because of many default libraries. python have multiple IDE platforms like Jupyter, Google colab, Pycharm and spyder. Python is used in top MNC's like IBM, Pixar, Netflix, Facebook and python is one of the main languages at Google, and YouTube is largely written in python. Even the most popular

independent research organization NASA uses it. This has high processing capacity and processing of big data became easy because of python. Machine learning, AI can be easily implemented in python.

B. Google Colab

Google colab is a brain child of Google research that is very much helpful for writing code in python. This will allow anyone to write a python code through browser. Google colab is different from other IDE's because this is browser based and this has the power to run code for machine learning, AI and data analysis. Google colab provides user with high GPU, CPU, TPU and supports most of machine learning libraries by default.

C. TensorFlow

TensorFlow is an machine learning and artificial intelligence library which is open source. This mainly focuses on training and inference of deep neural network. This can be used in programming languages like Java, JavaScript, C++ and mostly python. Programmers can use it to create large-scale neural networks with several layers. Classification, perception, understanding, discovering, prediction, and creation are some of the most common uses for Tensor Flow. Text-based apps, image identification, voice search, and many other applications use TensorFlow. TensorFlow is used for image identification in DeepFace, Facebook's image recognition technology. Apple's Siri uses it for voice recognition. We have used tensorflow for training the machine learning chatbot model in google colab.

D. NLTK

The Natural Language Toolkit (NLTK) is a Python programming environment for working with human language data in statistical natural language processing (NLP). It includes tokenization, parsing, categorization, stemming, tagging, and semantic reasoning text processing packages. We have used this toolkit for analysing the intent.json file and predict the answers for a new questions in the chatbot. Processing natural language became easy with the NLTK tool kit [9].

E. Pickle

Pickle is a python module that implements binary protocols for de-serializing and serializing. The process of converting python object hierarchy into byte stream is called "pickling" and inverse operation is called as "unpickling". pickle is used in machine learning so that training is not done multiple times. This keeps the track of serialized objects so that it won't be serialized again hence saves time and allows for faster execution. We used the pickle to save the model which is trained [8].

F. Tflearn

Tflearn is a highlevel abstraction layer under Tensorflow. Tflearn API is generally used for implementing deep learning neural networks. This is built on top of the Tensorflow and is used to speedup the experimentation's and still remaining compatible and transparent to Tensorflow. This is used to ease

and beautifully graph visualisations. This has built in neural networks which enables fast prototyping [8].

G. Numpy

Numpy is mainly used to work with multi-dimensional arrays and matrices. This has functions for working with linear algebra, matrices, and fourier transform. This is generally 30 times faster than python's general list. Numpy is homogeneous type and frees memory faster and arrays are densely packed. We have used this to create arrays and for bag of words techniques.

H. Lancaster Stemmer

Stemming is a process of reducing a word to its root word for information retrieval and extraction. Natural language processing uses Stemming to know the meaning of words. There might be many words with same meaning so this will reduce the word to root for finding actual meaning. This is mainly used in machine learning, AI, NLP, indexing systems, tagging and web search results to find the word we are searching for. There are two types of stemming Lancaster stemmer and porter stemmer. At the very basics of it, the major difference between the porter and lancaster stemming algorithms is that the lancaster stemmer is significantly more aggressive than the porter stemmer. The three major stemming algorithms in use today are Porter, Snowball(Porter2), and Lancaster (Paice-Husk), with the aggressiveness continuum basically following along those same lines. Porter is the least aggressive algorithm, with the specifics of each algorithm actually being fairly lengthy and technical. We have used lancaster stemmer to find root word from words list once we read it from intend.json file.

I. Bag of words

Bag of words is a text modeling technique under NLP. This technique is used in many areas of machine learning where NLP is needed. The main use of the bag of words technique that is used to pre-process a text by converting it into bag of words. The total number of instances of the most frequently used terms is kept in a bag of words. A table can be used to visualize this model, which shows the number of words that correspond to each term.

J. Google Cloud

Google Cloud is a cloud platform that provides user with resources that they cannot replicate in their actual location. This provides virtual access to a platform that is used by the google itself for hosting and building their websites and applications like Google search engine, YouTube, Gmail etc. Google cloud has high end CPU, GPU and TPU that support almost any kind of programs that will ease the processing speed for a program. We have used google cloud as our cloud platform for deploying our project using instances and micro-services and can be accessed anywhere at any point of time as long as the services are running.

K. Microservices

Microservices came into light in the late 90's and identified as important architecture that is currently in trend. This is used to create chunks of a large program and run them independently and hence making the microservice loosely coupled. This process is currently used not only in building new technology as well as in software teams. Two dependent development teams doesn't know what the part of development other team is doing and vice versa. Because of this there will not be direct dependency in the code and if there is a change in any part of the code in future other teams doesn't need to modify the code.

L. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

M. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as "3.5-inch disk drive".
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: "Wb/m²" or "webers per square meter", not "webers/m²". Spell out units when they appear in text: ". . . a few henries", not ". . . a few H".
- Use a zero before decimal points: "0.25", not ".25". Use "cm³", not "cc".)

N. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \quad (1)$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not "Eq. (1)" or "equation (1)", except at the beginning of a sentence: "Equation (1) is . . ."

O. *L^AT_EX*-Specific Advice

Please use “soft” (e.g., `\eqref{Eq}`) cross references instead of “hard” references (e.g., (1)). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don’t use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in *L^AT_EX* will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you’ve discovered a new method of counting.

BIB_TE_X does not work by magic. It doesn’t get the bibliographic data from thin air but from .bib files. If you use *BIB_TE_X* to produce a bibliography you must send the .bib files.

L^AT_EX can’t read your mind. If you assign the same label to a subsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

L^AT_EX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it’s supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won’t be any anyway) and it might stop a wanted equation number in the surrounding equation.

P. *Some Common Mistakes*

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum μ_0 , and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
- Do not use the word “essentially” to mean “approximately” or “effectively”.
- In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.

- Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
- Do not confuse “imply” and “infer”.
- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the “et” in the Latin abbreviation “et al.”.
- The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [5].

Q. *Authors and Affiliations*

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

R. *Identify the Headings*

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

S. *Figures and Tables*

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an

TABLE I
TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	Subhead	Subhead
copy	More table copy ^a		

^aSample of a Table footnote.



Fig. 1. Example of a figure caption.

example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [3]. Papers that have been accepted for publication should be cited as “in press” [4]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [4].

REFERENCES

- [1] <https://en.wikipedia.org/wiki/Chatbot>
- [2] https://en.wikipedia.org/wiki/Natural_Language_Toolkit https://en.wikipedia.org/wiki/Natural_language_processing

- [3] https://en.wikipedia.org/wiki/Machine_learning <https://en.wikipedia.org/wiki/TensorFlow>
- [4] https://en.wikipedia.org/wiki/Google_Cloud_Platform https://en.wikipedia.org/wiki/Bag_of_words_model
- [5] <https://en.wikipedia.org/wiki/Microservices>
- [6] <https://en.wikipedia.org/wiki/PuTTY>
- [7] <https://www.codeproject.com/Articles/1204813/Lina-ChatBot-Generating-Response-Using-Docment-Re>
- [8] Kandpal, Prathamesh, et al. “Contextual Chatbot for healthcare purposes (using deep learning).” 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4). IEEE, 2020.
- [9] Shah, S., and Seema Shah. “A comparison of various chatbot frameworks.” J. Multi-Criteria Decis. Anal 6 (2019): 375-383.
- [10] Ciaburro, Giuseppe, V. Kishore Ayyadevara, and Alexis Perrier. Hands-on machine learning on google cloud platform: Implementing smart and efficient analytics using cloud ml engine. Packt Publishing Ltd, 2018.
- [11] Hakkarainen, Jarmo. “Three major cloud providers and their applicability for hosting a chatbot.” (2017).

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.