IMAGIBOT – An Image Recognition Chatbot for Sri Lankan Ancient Places

P J Senevirathne# and D U Vidanagama

General Sir John Kotelawala Defence University, Sri Lanka

#pabodasenevirathne@gmail.com

Abstract: Ancient artifacts of Sri Lanka are used as major sources of getting historical information about Sri Lanka. Acquiring the necessary information about those artifacts become a huge challenge for the visitant of artifacts. Text based search engines are typically used to retrieve information about ancient artifacts of Sri Lanka. These systems require the user to formulate a text query that provides information such as the place, where the object is placed or where does that can be found, what is the century object belonged. Oppositely visual search systems can be used. They provide information to users of the system such as scholars, tourists, local explores in a most intuitive and immediate way by using an image as a query. This research involves in developing a chat application along with Convolutional Neural Networks for image recognition of ancient artifacts. The image recognition model will be a part of the chat bot that has ability to retrieve more information about recognized images. Convolutional Neural Networks, Recurrent Neural Network, TensorFlow, Keras have been used as core technologies in this research project. As the research involves in developing a chat application with image recognition capabilities, the application will improve the knowledge sharing of Sri Lankan ancient legacy to the word in an effective way.

Keywords: Chatbot, Convolutional Neural Networks, Ancient Artifacts, LSTM

Introduction

The sacred places like Anuradhapura, Polonnaruwa and Kandy in Sri Lanka, taking more attraction among tourists from different countries because of rich diversity of ancient Sri Lankan culture and architecture. Those can be considered as major centres of Sri Lankan civilization.

Sri Lankan history is started in 5th century B.C with the first settlement that make by Indo Aryan who came from north India in this land. After that Sri Lanka had a continuous settlement and Sri Lanka has continued records about all those things too. That is the reason that Sri Lanka has a great documented history. Those things have added a great value to Sri Lanka.

As many peoples tending to visit ancient places in Sri Lanka with a high value in order to find out historical information for their research purposes, there are some difficulties that they are facing when gather information about those ancient things. Lack of organization of description demonstrations, reading long historical descriptions that has kept in front of ancient artifacts in a busy time schedules in a limited time periods, Sometimes, having some images of ancient things that don't have its identification and related information are some of them. In such situations, sometimes they have to move around more web sites and historical sources, in order to gather historical information.

Technology related with machine learning, image recognition has become a mandatory

thing in worldwide and those technologies has already applied in travel and education industries. But in Sri Lanka a there are not many automated systems that deal with Artificial intelligence in image recognition.

To make better knowledge sharing among tourists and local people about ancient legacy of Sri Lanka, an intelligent chatbot (ImagIBot) system can be designed and developed. The system will introduce a chatbot as a human-computer dialog system to communicate between user and device by using both image and text as a query.

By using this intelligent chatbot system called ImagIBot, has image recognition capabilities and users can chat with a real time bot more efficiently to gain knowledge about those historical artifacts. For that, system will allow users to insert a photo of a historical place or artifacts easily to the chatbot, which they have. Thought the chatbot user can easily ask what they want to know about the artifact in the image, instead of reading all details of them by searching through many web sites, books or in other historical sources.

System will help to all foreign and local travellers to find out places more efficiently through the chatbot, and this will be a real time experience that they can gain by the chatting with a chatbot. It will be able to give accurate information and locations on a particular historical heritage and save the time of users by providing necessary information according to their information requirements by eliminating reading long stories and historical information searching by moving into many web sites.

This will be a good experience with improved user service, better engagement, keeping up with the trends by being present on messaging platforms, user satisfaction and time savings.

Apart from those the chat bot have the ability to do chat escalation to a human agent when

necessary, store user queries that it cannot identified in real time chat for future training, spelling correction abilities, ability to upload images easily. The information that the system provides will make user to work efficiently and leads to enhancement in performance in their searching purposes with more efficiency. As business verticals, the system can be successfully used in Travel and Education industries.

Through the paper will present design and implementation of the system. Section 2 of the paper describes overview of some existing chatbot systems. Section 3 continues the Methodology of the system; section 4 continues with the evaluation and section 5 continues with the discussion. Finally, Section 6 draws together a conclusion of the work done and discusses future research directions.

Literature Review

Chatbot is a program designed counterfeit a smart communication on a text or spoken ways. Those chatbots are widely popular in the world. And they are responding intelligently like human. (Dahiya , 2017). Those give an automatic reply to human what they are looking for. Not only that chatbots has capability to interact with users using natural language processing. Because of that that chatbots can be used as a tool of entertainment such as pandorabots and ELIZA. In this way chatbots can be used as a,

Tool of entertainment

Chatbots can be used as a tool to learn and practice a language.

It can be used as an information retrieval tool.

Can be used as assistants in E-Commerce, business and in other domains. (Shawar & Atwell, 2007)

Today there are many conversational agents (CA) such as Siri, Google Assistant and Alexa. Not only that Cortana, Pandrabots, Burberry are found on Messenger, Slack, or in Skype.

The first chatbot was emerged in 1996 from MIT. That is called ELIZA that worked on simple declarative rules. The follow up chatbots include ALICE, Elizabeth, Loebner winners such as Mitsuku, Rose etc. Artificial Linguistic Internet Computer Entity (ALICE) is an award winning chatbot that available as open source that has developed using AIML (artificial intelligence Mark-up Language) (Jain, et al., 2018). And this was developed by Dr. Wallace and ALICE uses simple pattern machine algorithm. (Bani & Singh, 2017)

The GUS stands for the Genial Understander System, can be describe as the first experimental computer system that did in language understanding. It is also able to conduct a more or less realistic dialogs within some limitations which is a frame driven dialog system. Figure 1, shows a transcript of an actual dialog of GUS. There are some problems of the natural dialogues such as,

  i. Mixed initiative

  ii. Indirect answers

Figure 25: A transcript of an actual dialog with GUS
Source: GUS, A frame driven dialogue system (1976)

It uses Frames that can be used in reasoning component and they can used to represent collections of information at many levels within their system. Not only that, frame is a data structure that potentially contain a name, reference to a prototype frame and set of slots. Apart from that it follows a simple depth-first, recursive process. That means it

systematically completing work on a given slot before continuing to the next. This is the way of GUS attempts to retain the initiative in the dialogue. GUS is not very intelligent itself, but it illustrates what that is believed to be essential components of such system. (Bobrow, et al., 1976)

The research paper written by Vladimir Vlasov represented a dialogue policy based on transformer architecture and compares the performance of Transformer embedding Dialogue that knows as TED policy to LSTM and to REDP in order to overcome the limitations of RNN. Recurrent neural networks (RNN) are used in both open domains and task-oriented systems in order to process the sequence of previous dialogue turns in recent years. (Vlasov, et al., 2019)

When designing a chatbot, open source languages such as AIML and Chat Script can be used. They can be introduced as the mostly used open source languages that used to design chatbots. AIML is an XML based mark-up language and it is the most used chatbot language because of its simplicity. Chat Script is a scripting language and it can accept user text input and generate a text response that help to manipulating natural language. (Arsovski & Muniru, 2017)

There are two types of dialogue systems. They are,

  i. Goal driven systems – Ex: support services

  ii. Non goal driven systems – Ex: language learning tools/computer game characters.

Slot filling techniques can be use in rule-based approach such as Siri that proven to be reliable. Because in rule base approach chatbot answers questions based on rules on which is trained on. But in some cases, bot is not efficient in answering questions when pattern is not match with the rules of the bot that has trained. The purpose of that is the

bot that functioning efficient than rule base approach. Because end to end trainable memory can enhanced neural networks architecture. (Thomas & Thomas, 2018)

As museums and galleries have track records, it is experimenting new ways to reach their audience with the use of emergent technologies. Emerging free chatbot creating platforms such as chatfuel, chatterbot, Eliza enable opportunity to be experimenting with chatbots with low effort, low cost and keeping staff resources at a low level. There are many numbers of museums that using bots to engage their audience with new technologies. Some of museums that uses chatbots are listed below,

    i.   Heinz Nixdorf Museums Forum - The Heinz Nixdorf MuseumsForum in Paderborn Germany has an early experience of using an avator bot introduced as MAX. Developed in 2004, It directly engages with visitors through a screen as a virtual museum guide.

    ii.   San Francisco Museum of Modern Art (SFMOMA) - Send Me SFMOMA is an SMS service that provides an approachable, personal, and creative method of sharing the breadth of SFMOMA's collection with the public (Boiano, et al., 2018)

CNN is a most impressive forms of ANN (Artificial neural networks) that primarily use to solve difficult image driven pattern recognition. Artificial Neural Networks (ANN) are mainly comprised of high number of interconnected computational nodes. Those have two types of learning methods such as,

    i.   Supervised learning - supervised learning is the learning through pre labelled inputs, which act as targets.

    ii.   Unsupervised learning - unsupervised earning is different

form it does not include any labels for learning.

The only different between convolutional neural networks and artificial networks are CNN is mostly used in field of pattern recognition within images. (O'Shea & Nash, 2015)

Large deep convolutional neural networks can have good results in image classification, but those image classification needs methods to reduce and prevent the over fitting. In past neural networks are not used in Computer vision because it needed more computational power in order to train and need lots of labeled data too. But today fast GPUs and big labeled image datasets can be used in order to solve these problems. (Smirnov, et al., 2014). Figure 2 shows the architecture of a deep convolutional neural network.
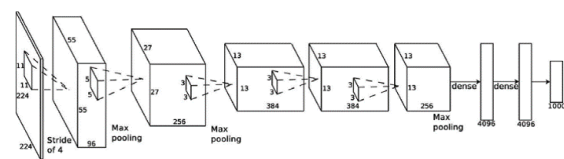

Figure 26: Deep convolutional neural network
Source: ImageNet (2014)

There are traditional machine learning methods like multilayer perception machines, support vector machine. But today CNN has become more powerful and universal deep leaning mode. Because CNN can be identified as a multilayer neural network, it is a most classical and common deep learning framework and it has excellent performance in field of hyperspectral image classification. Because of all those things convolutional neural networks has been more effective in image processing when considering traditional machine learning methods. (Xin & Wang, 2019)

Methodology

The section describes the methodology of the system. This research project aims to build an image recognition model in order to

classify ancient images of Sri Lanka using convolutional neural networks and combine it with a chatterbot. This will provide best quality information to user.

The inputs for this automation process are images and texts. First user will be uploading the historical images to the chat interface and through the image recognition model developed using CNN will recognize the ancient artifact. After that it will send the related name to the user through the chat interface. Then user can chat with the ImagIBot and ask more related information about those ancient artifacts of Sri Lanka.

Furthermore, chatbot system can do some small operations such as have the ability to do chat escalation to a human agent when necessary, ability to upload images, store user queries that it cannot classify in real time by the chat modal (finally, those queries uses for future training), spelling correction abilities etc.
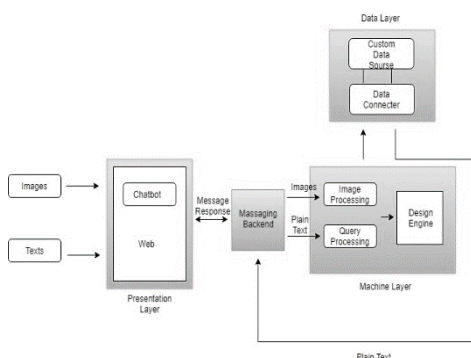

Figure 27. Overall ImagIBot system architecture
Source: Author

As programming methodology, Extreme programming has been used, because it is an Agile software development framework that helps to provide high quality software. The main reason for that is Extreme programming helps to dynamically changing requirements.

The overall system architecture is lies between three main architectural layers that can be called,

  i.   Presentation layer

  ii.   Application layer

  iii.   Database layer

When look by the architectural perspectives that shows in figure 3 , it is very critical task to identify and acknowledge the workflow of each layer in order to design the system accordingly and correctly.

A. Presentation Layer

Presentation layer is the layer where the user directly interacting with the ImagIBot. Front end of the system includes all the interfaces of the system. HTML5, CSS3, JavaScript, Bootstrap along with the Angular Framework has been as front end technologies.

B. Application Layer

Application layer, if not the machine learning layer of the proposed system is designed with the two main modules. This is the layer which all the image classification happens. This layer will execute all the main goals of the system. The two separate modules in this application layer are named as,

  i.   Chat module

  ii.   Image recognition module

1) Chat Module:

The chat module is the part that handles all the user queries. If query is a text, the chat module takes text as the input and it will understand the intent of the sentence. If query is an image the chat module will send it to the image classification module that have developed using CNN. When user enter the query in text format, it first classifies the intent of the user query and then extract entities such as artifact place, opening hours, creator, only if entities are visible in the user query. The GUS (Genial Understander System) is used as the concept behind the chatbot. As GUS is a frame driven dialog system. GUS has ability to conduct a more or less realistic dialogs within some limitations. The Frames are used in the chatbot module because frame is a data structure that

potentially contain a name, reference to a prototype frame and set of slots. (Bobrow, et al., 1976).

As an Example ,When a user asks, "What is Sigiriya",the intent will be "asking about an ancient artifact" and the entity will be the "Sigiriya". Internally the model uses the bag of word (Bow) algorithm to find the intent of the user and Conditional Random Field (CRF) to find entities of the user inputs. A python library known as pyspellchecker has used to create a custom dictionary for the ImagIBot in order to correct spelling of user text queries before classifying the intent. Otherwise meanings of the text queries can be change because of misspellings.

According to chat bot architecture presented, user has ability to enter query in both text and the image format. When user insert an image to the ImagIBot, it will classify the image using image recognition module and it will send the name of the artifact in the image to the chatbot module. After that the artifact name will be saved in a slot of a chatbot module for further usage. Finally, user can ask more information from the chatbot.
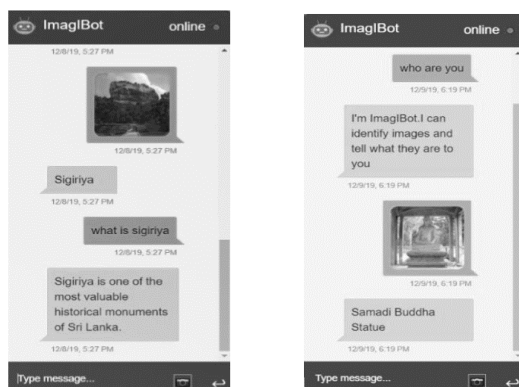

Figure 28: Successful image recognition and some conversations of the ImagIBot

When chatbot need some information from database, it calls the database connection according to the slot that filled with artifact name and retrieves necessary data from database. When making utterances those data will be used. Python Flask has been used for connecting components tighter in the ImagIBot. As initial step, Rasa (A open source

machine learning framework) has been used to develop the AI assistant. (Vlasov, et al., 2019)

2) Image Recognition Module:

Image recognition model recognizes all the images that user uploaded to the ImagIBot. It has ability to identify images only belongs to the Sri Lankan ancient legacy. It is totally based on a model that developed using Convolutional Neural Networks. As a class in deep learning neural network, convolutional neural networks can be identified. The image recognition model of the ImagIBot has been developed using CNN, because it has ability to differentiate images one from the other. Not only that CNN model will have capability to send fallback message to user, if the probability of being a Sri Lankan artifact is low or if it is not in the ImagIBot's domain. Steps used to develop CNN model is as follows,

Convolutional Layer – The convolutional layer of the system will make use of a set of learnable filters in convolutional neural networks.

Pooling Layer - This layer is located between convolutional layers in CNN architecture. This reduces the number of parameters and computation in the network. And most importantly, it controls the overfitting. This is done by progressively reducing the spatial size of the network.

Max pooling – Max pooling has used as a noise suppressant because it takes out only the maximum form the pool. This layer is also responsible in order to reduce spatial size of the convolved feature.

Flattening – This is a very important layer in the image recognition model because it is used to convert images in to 1 dimensional array that is used for the inputting that to next layer. This create a single long feature vector. Apart from that this layer is

connected to the last classification model that is known as fully connected layer.

Fully Connected Layer - In fully connected layer all neurons have complete connection to all the activations from the previous layers.

Apart from that 'categorical_crossentropy' has used for the multi class classification in image recognition. 'relu' and 'softmax' has been used as activation functions. In this way Convolutional neural network are made up of hidden layers and the fully connected layers. TensorFlow has been used in CNN in following purposes, some of them are listed below,

  i.   Preprocess the images that is in the dataset

  ii.   Create placeholders in the model

  iii. In order to Combine all functions into a model

Keras has designed to enable to do fast experimentation with deep neural networks, When developing CNN model,

  i.   Keras library is very helpful in making that model in a simple way

  ii.   Keras is very useful in training the model in better way

  iii. Used for data preprocessing

  iv.   Most importantly, it has been Used ImageDataGenerator class for generate batches of tensor images with the real time data augmentation in the CNN model.

C. Database Layer

This layer is all about managing the database of the system developed using MySQL database. Because of that this layer is responsible for managing all database requirements of the system.

Evaluation

The system can be evaluate using following measures. The system has good comprehension capabilities, because ImagIBot ensures good texting and error free experience of the user by using 'spelling correction' feature. This helps to improve the accuracy of the intent classification of text and give most meaningful reply to the user. As the system has rich data set and use of best hyper parameters such as activation functions, best number of convolutional layers and best number of nodes in each convolutional layer will increases the performance of the image recognition model. Not only that ImagIBot has good user engagement, because it has capability to initiating conversations with users and interact with them in order to share information with users more effectively. ImagIBot has programmed with capability to fetch image and text query quickly and respond quickly. Because of that the system will respond to the user quickly and will increase the user satisfaction. The accuracy of the image classification model has been calculated as follow,

True Positive(TP)	False Positive(FP)
Number of TP results: 1	Number of FP results: 1
False Negative(FN)	True Negative(TN)
Number of FN results: 8	Number of TN results: 90

Figure 5: Accuracy of the Image Classification model of the ImagIBot
Source: Author

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{1 + 90}{1 + 90 + 1 + 8} = 0.91$$

The 0.91 or 91% accuracy comes out from the image recognition model outputs 91 correct predictions out of 100 total examples. Not only that the system will have may functionalities and it can be scalable for additional modules and other domains

easily. Finally, ImagIBot has used rich media images and text in order to get user's attention.

Discusssion

ImagIBot is a chatbot that have image recognition capabilities. It has been developed with the trending artificial intelligent algorithms such as convolutional neural networks, Bow, CRF and Recurrent neural networks. The system is gives ability to user to upload images that belongs to the ancient Sri Lanka. After that it will identify all those images and tell the user what is that artifact is. At the end of that process user can chat with the real time bot developed and get knowledge about those ancient artifacts.

The system is helpful to many people such as foreign tourists, school children, teachers, local travelers, and for ancient artifact explorers, when they search information about Sri Lankan ancient history by visiting some museums, ancient places, templets etc. They can easily take a photograph of an artifact and upload it to ImagIBot in order to find information about particular artifacts. It will be a good solution for problems that occurs in those places, when gather information about ancient artifacts. The system has ability to give accurate information about historical heritage also. Because of that ImagIBot saves the time of users by providing important information according to their information requirements by eliminating reading long stories and historical information searching by moving into many web sites. Because of that users can use this Intelligent system for gain knowledge and find important information about Sri Lankan ancient artifacts and places.

Finally, this system will be a valuable and efficient system to the people those who visit and the search about ancient legacy of Sri Lanka. Not only that information that the system provides will make user to work efficiently and leads to enhancement in performance in their searching purposes of ancient historical artifacts of Sri Lanka. The system can implement in Sri Lankan museums as part of their audience engagement program.

Conclusion and Further Work

This paper presented design and implementation of a Chatbot system, which is designed as a human-computer dialog system. Through the development process start to end, it has been successful enough to meet main requirements of the image recognition chatbot in order to recognize ancient places and artifacts. This will helpful to share ancient legacy of Sri Lanka more effective way using Artificial Intelligence. Not only that, As business verticals, the system can be used in Travel and Education industries world widely. But it does not mean that the system cannot improve it domain, technologies and new functions further.

Today technologies are growing faster and with the growing of those technologies there can be done many enhancements to the ImagIBot. Some of them have listed below that can be done within a short period of time with the technology enhancements. Add Context switching capabilities, Integrations with the Alexa, add multilingual supports like Sinhala language are some of them. Based on chat history sentimental analysis can be done regarding user satisfied or not by chatting with the ImagIBot. Not only that the ImagIBot can be integrated to other channels such as WhatsApp, Facebook Messenger or in Slack. The process of the chatbot system will be much transparent to the user. That will be easier to incorporate any development in early states of the use of the chatbot system.

References

Arsovski, S. & Muniru, I., 2017. ANALYSIS OF THE CHATBOT OPEN SOURCE LANGUAGES AIML AND CHATSCRIPT:A REVIEW. p. 8.

Bani, B. S. & Singh, A. P., 2017. College Enquiry Chatbot Using A.L.I.C.E. International Journal of New Technology and Research (IJNTR), 3(1), p. 2.

Bobrow, D. G. et al., 1976. GUS, A Frame-Driven Dia|og System. p. 19.

Boiano, S. et al., 2018. Chatbots in museums: hype or opportunity?. p. 12.

Dahiya , M., 2017. A Tool of Conversation: Chatbot. INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING , 5(5), p. 5.

Jain, M., Kumar, P., Kota, R. & Patel, S. N., 2018. Evaluating and Informing the Design of Chatbots. p. 12.

O'Shea, K. & Nash, R., 2015. An Introduction to Convolutional Neural Networks. p. 11.

Shawar, B. A. & Atwell, E., 2007. Chatbots: Are they Really Useful?. p. 22.

Smirnov, E. A., Timoshenko, D. M. & Andrianov, S. N., 2014. Comparison of Regularization Methods for ImageNet Classification with Deep Convolutional Neural Networks. s.l., s.n., p. 6.

Thomas, J. S. & Thomas, S., 2018. Chatbot Using Gated End-to-End Memory Networks. International Research Journal of Engineering and Technology (IRJET), 5(3), p. 6.

Vlasov, V., Mosig, J. E. M. & Nichol, A., 2019. Dialogue Transformers. p. 7.

Xin, M. & Wang, Y., 2019. Research on image classification model. EURASIP Journal on Image, p. 11.

Acknowledgment