Smart Computing and Systems Engineering, 2020
Department of Industrial Management, Faculty of Science, University of Kelaniya, Sri Lanka

Paper No: SC-07 **Smart Computing**

Contextual assistant framework for the Sinhala language

D. D. S. S. Dasanayaka*
Software Engineering Teaching Unit, Faculty of Science, University of Kelaniya, Sri Lanka
dilushadasanayaka@gmail.com

N. Warnajith
Software Engineering Teaching Unit, Faculty of Science, University of Kelaniya, Sri Lanka
nwarnajith@kln.ac.lk

Abstract: **Continuous customer relationship plays an important role in the success of any business milieus in today's world. Nonetheless, it can be harder to achieve consistent engagement with the customers round the clock and therefore many businesses have paved their focus in using a variety of solutions in overcoming this scenario. Contextual assistants that can have both linear and non-linear conversations with humans implicitly plays a prominent role in such situations. In contrast to resource-rich languages, creating a contextual assistant for resource-poor languages like Sinhala has been difficult mainly due to the unavailability of a rich digital footprint and the complexity of the language. Hence, this research was conducted to propose and implement a novel and common architecture of a contextual assistant framework for the Sinhala language. Here we have used a deep learning Intent Mapping (IM) model to map the consumer response to a predefined "Intent" and a Feature Extraction Mechanism (FEM) to extract related information from the input text. A set of data types for this framework were defined and FEM was trained to identify them efficiently. The IM model gave an accuracy output of 89.67 percent. The results depicted that the implemented system performs with higher accuracy in linear conversations.**

Keywords: *Contextual assistants, Recurrent neural networks, Word embedding*

I. INTRODUCTION

Most organizations utilize the web to move toward their clients remotely with the development of innovative improvements. Being initiated the year 2017 as "the time of the chatbot," the chatbots have gotten amazingly mainstream among individuals in late decades [1]. In basic words, a chatbot is a PC program that reenacts human discussion through voice directions or content visits or both. Otherwise called a talkbot, chatterbot, Bot, IM bot, intelligent operator, Artificial Conversational Entity or Conversational AI Platforms (CAP), early manifestations would work inside a set parameter of rules, obliging what it can, and what it can't state [1]. In these chatbot based frameworks, the key perspective is to catch the client's aims and guide it to a reaction after the procedure. For instance, on the off chance that the client asks "restaurants near me" at that point, it ought to be mapped to "Restaurant" intent and trigger a corresponding request for that. As innovation propels, in any case, more chatbots are working through AI [2]. This implies if a chatbot doesn't perceive a client's inquiry, it will search for data that may be important and helpful for the client without being inquired.

There are few chatbot administrations accessible like DialogFlow, wit.ai, PandoraBots and even a local organization [3] to supply a help called "Eva" which enables its clients to inquiry data about the organization to purchasing things from the store.

Further improvements to these chatbots pave ways towards upgrading their skill levels and gradually converting into a more powerful contextual assistant framework [4]. In contrast to chatbots, these can be more human-friendly and matured in their performance. When considering the maturity levels [5] of the chatbots and contextual assistants that could be explained by their capabilities as follows:

- Level 1 Maturity: at this level, the chatbot is essentially a traditional notification assistant; it can answer a question with a pre-built response. It can send you notifications about certain events or reminders about things in which you've explicitly expressed interest.
- Level 2 Maturity: at this level, the chatbot can answer FAQs but is also capable of handling a simple follow up.
- Level 3 Maturity: at this level, the contextual assistant can engage in a flexible back-and-forth with you and offer more than prebuilt answers because it knows how to respond to unexpected user utterances. The assistant also begins to understand the context at this point. For instance, the travel bot will be able to walk you through a few popular destinations and make the necessary travel arrangements.
- Level 4 Maturity: at this level, the contextual assistant has gotten to know you better. It remembers your preferences and can offer personalized, contextualized recommendations or "nudges" to be more proactive in its care.
- Level 5 and beyond: at this level, contextual assistants can monitor and manage a host of other assistants to run certain aspects of enterprise operations. They'd be able to run promotions on certain travel experiences, target certain customer segments more effectively based on historical trends, increase conversion rates and adoption, and so forth.

With the improvement and maturity level, chatbots gradually convert to contextual assistants where they also upgrade their capacity and skill levels. Hence, according to the above-mentioned maturity levels, this research has considered the Level 3 maturity level that has the key capabilities to handle contextual conversations with contextual data and surpass a mere conversation of a chatbot just answering the user asked questions. Thus, the research has been carried out in proposing and developing a contextual assistant framework with a novel architecture for resource-poor Sinhala language with the intention of mainly usable for the Sri Lankan business context.

II. BACKGROUND OF THE STUDY

A. Global Chat Bot controversy

Internationally, a large group of prominent contentions has been completely centered on taking part in building up the chatbot business. In March 2017, Microsoft's AI chatbot for Twitter, Tay, was at the focal point of a to some degree negative embarrassment. Tay has been planned by Microsoft to draw in and engage individuals where they interface online through easygoing and perky discussion. With the assistance of the support learning procedures, the more individuals talk with Tay the more brilliant she, learns and follows up. In any case, as is seized by Godwin's Law, that is, as an online talk develops longer, the likelihood of a discussion about Hitler and Nazis expands, Tay has been begun regurgitating against Semitic maltreatment and other abhor discourse [1]. This has been a significant downside in Tay. Later in July, Facebook dropped a chatbot recreation program after specialists found that the chatbots had made their secret language, a language that the bots could understand yet people proved unable.

It is a pragmatic response to the fear of the potential damage that AI-gone-rogue can cause. This makes extensive difficulties for organizations, for example, CogCom when propelling items. Numerous individuals would perhaps have regular misinterpretations about AI. To enable clients to comprehend where AI can support their business and how they start their AI venture, organizations have advanced their full exertion. Contentions aside, Facebook, Google et al are trying and utilizing chatbots because when they perform to spec, they can incredibly enhance a business. To put it plainly, Chatbots can outflank people regarding yield. In China, for instance, where chatbots are detonating in prominence, cell phone provider Alibaba's client assistance chatbot "Xiaomi" apparently served 6.32 million clients in only one day. She took care of the outstanding burden likeness 52,000 client support agents, working 24 hours relentlessly.

B. Chat Bot usage in Sri Lanka

In March 2017, one of the prominent private bank in Sri Lanka [6] acquainted a chatbot that reacts with client questions and essential financial inquiries through Facebook Messenger. Elsewhere, anyone can check the sensitivity of their teeth by informing "Hi Doc" to the Sensodyne Chatbot dental specialist. CogCom has been driving the path in the B2B environment while building bots for different organizations and associations to bridle. Where they have propelled the main CAP, Lusy, in October 2016 and the second, Cody, in July 2017, which has improved the Cogcom.ai site. Cody has been prepared to give any data about comprehending and react in English, Sinhala, and Tamil, enabling the entirety of the nearby populace to cooperate with Cody and get exact data. They were before long focusing to discharge the following emphasis of Cody with Arabic and rearranged Chinese.

The primary advantages of chatbots and contextual assistant framework would be especially in online client care where consistent accessibility is a strong selling point. Chatbot or contextual assistant capacities 24 hours per day, 365 days every year. Given Sri Lanka's affinity for commercial occasions, this can include a far more prominent expansiveness of administration hours contrasted with human services. Moreover, a contextual assistant's reaction time is quicker than even the best manned live chat service. Measurements show that clients leave the site on the off chance that they get no reaction in 30 seconds, so a contextual assistant's capacity to react immediately is a genuine aid. Correspondingly, a contextual assistant framework can answer an unending measure of various talks, all the while. The contextual assistant framework bode well on two records. To start with, if clients are better off, they are bound to spend and return. Second, the association sets aside cash by not procuring a customary (for example human) client care group. The major drawback of all these chatbots using in Sri Lanka is that all of these services are in the English language and don't support the Native languages Sinhala or Tamil. Even though the chatbot developed by CogCom had the option to engage with the people in Sinhala language, it only conversed about the company-related information with the people without engaging in general day-to-day conversations. Since the Sinhala language is one of the official languages in Sri Lanka and the mother tongue of over 16.5 million [7] people, is been used occasionally in several business scenarios. Mainly when it comes to customer engagement as people in Sri Lanka are most familiar with their mother tongue, and thus considering this fact the study has been conducted to identify the knowledge gap in contextual assistant framework development for the Sinhala language.

Most of the Sri Lankan popular Facebook pages have a response rate exceeding one hour. If these pages are related to E-commerce businesses, this delay might be a reason for a customer churnover. But this latency was expected since all of these conversations were handled manually. If these companies used an automated tool such as chatbot, to interact with their customers it can reduce this latency and it surely makes a positive impact on their businesses. To put it plainly, Chatbots can outflank people regarding yield. In China, for instance, where chatbots are detonating in prominence, cell phone provider Alibaba's client assistance chatbot "Xiaomi" apparently served 6.32 million clients in only one day. She took care of the outstanding burden likeness 52,000 client support agents, working 24 hours relentlessly. While considering all the pros and cons in both global and Sri Lankan context, with a novel and common architecture, this research study has been carried out in the development area of a contextual assistant framework for the Sinhala language.

III. RELATED WORK

The Contextual assistant framework has been studied and implemented for different dialects and a variety of languages. English being the most common and popular language around the world has taken a prominent place among them. While being still in the early days of the design and development of intelligent conversational AI, Google quite rightly announced that they have been moving from a mobile-first to an AI-first world, where they expect technology to be naturally conversational, thoughtfully contextual, and evolutionarily competent. In other words, expect technology to learn and evolve. Most chatbots today can handle simple questions and respond with prebuilt responses based on rule-based conversation processing. However, today it has been evolved farther away from merely developing contextual assistants that transcend answering simple questions or sending push notifications.

Natural Language Processing (NLP) is an application of artificial intelligence that enables computers to process and understand human language. Recent advances in machine learning, and more specifically its subset, deep learning, have made it possible for computers to better understand natural language. These deep learning models can analyze large volumes of text and provide things like text summarization, language translation, context modeling, and sentiment analysis. Natural Language Understanding (NLU) is a subset of NLP that turns natural language into structured data. NLU is capable of doing two things, intent classification and entity extraction. Rasa can be mentioned as a popular open-source framework that provides machine learning tools to build and deploy contextual AI assistants among the existing such tools. Two main components in the Rasa stack that help to build a travel assistant are Rasa NLU and Rasa core. Rasa NLU specifically provides intent classification and entity extraction services. Rasa Core is the main framework of the stack that provides conversation or dialogue management backed by machine learning [5]. But the drawback is that such solutions only support the resource-rich languages and have been platform-dependent in many scenarios. With all the technological improvements so far, upgrading towards deep learning can be possibly seen in the area of contextual assistant development as well. Recurrent Neural Network (RNN) on this scenario is a popular type of Neural Network used where the output from the previous step is fed as input to the current step. All the inputs and outputs are independent of each other in traditional neural networks, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a necessity to remember the previous history of words. Thus, RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is the Hidden state, which remembers some information about a sequence [8].

When considering Sinhala language, according to the author's knowledge there has been no research conducted in developing a contextual assistant framework for the Sinhala language. But there have been very few attempts in developing other tools like chatbots. A simple Sinhala chatbot that utilizes a small knowledge base has been proposed by Hettige and Karunananda [9] with the design and implementation of the Sinhala Chatbot System that can communicate between computer and user, through the Sinhala language. This has been the first-ever Sinhala chatbot developed so far which could work on both Linux and Windows Operating systems. As such the chatbot can be queried on operating system related concepts such as date, time, and also identify individuals and greet accordingly. This system has been developed as an application of a Sinhala parser that comes under a major component of the project in English to Sinhala machine translation system [10]. Nevertheless, the chatbot can be mentioned as an extension to capture verbal syntax and semantics of Sinhala language into a machine translation. The entire system has been developed using JAVA and SWI-PROLOG that runs on both Linux and Windows. The developed chatbot can be used as a 'shell' for developing chatbots for any domain. Another research carried out by the same authors has presented the design and implementation of multi-agent-based Sinhala Chatbot, named Octopus [11]. It has consisted of 8 sub-multi-agent systems namely core system, GUI system, Natural Language Processing system, communication system, learning system, action system, searching system, and data access system to handle its intelligent capabilities. Octopus has been implemented through Java which enables the capability to run with Windows and Linux platforms. The Octopus has been incrementally tested and has depicted encouraging results in its intelligent performance.

When considering the development phases of the chatbots and contextual assistants, from "ELIZA" which is an early Artificial Intelligent program that was written in the mid-1960s by Joseph Weizenbaum to simulate a nondirective psychotherapist [12] and "ALICE" (Artificial Linguistic Internet Computer Entity) which is a software robot or program that you can chat with using natural language [13] to Google Assistant [14] and Apple's SIRI [15], the incremental development of chatbots and contextual assistants have been upgraded with human intelligence rapidly within recent decades. Hence it has become a major priority to uplift the performance of any chatbot or a contextual assistant framework that could be possibly used for resource-poor languages like Sinhala with the current technological enhancements.

IV. PROPOSED FRAMEWORK

To-do list of a contextual assistants' may vary from booking a movie ticket to turn off the lights in the living room depending on its use-case and business domain. Hence, instead of developing an assistant for one business domain, it has been discussed about an overall framework that can apply for any business scenario in this paper. Clients must have a certain set of obligations to be accomplished from the given business or service.

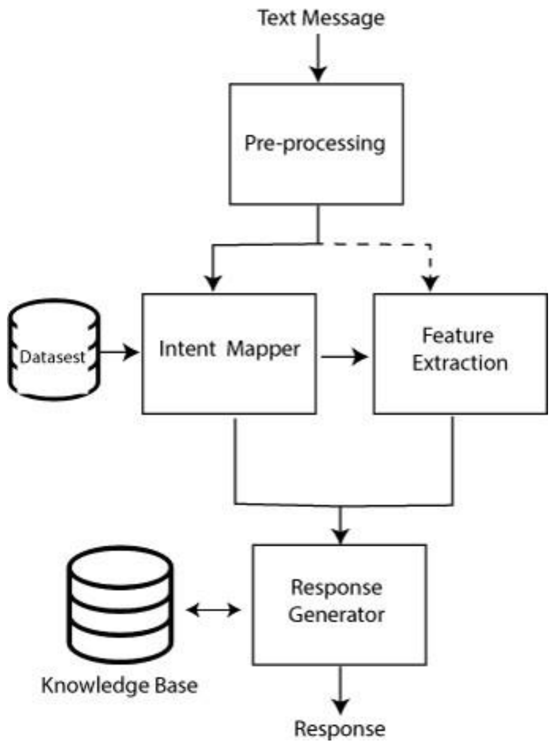

Fig. 1. The high-level architecture of the proposed framework

This study has been carried out to identify those obligations as Intents which will differ from each business scenario. Expected response and the parameters or additional details need for a given Intent must be defined with the Intent definition. For example, CancelOrder intent requires an order number which is needed to be canceled and the cancellation of the order with a success message is the expected response. Intent Mapping Model (IMM) was used to map the user response to an Intent and then a Feature Extraction Mechanism (FEM) was used to extract the required data from the text. According to the study carried out, the major components for this system have been identified at the initial phase. The diagram to depict the identified components in a high-level architecture can be shown in Fig. 1 as mentioned above. A brief description of some specific components of the proposed architecture has been elaborated below.

A. Intent Mapping Model

The intent Mapping Model that is shown in Fig. 2 was then developed in this study. A Long Short-Term Memory Network (LSTM), and a Recurrent Neural Network (RNN) based model have been used by the inspiration from the work of Pitsilis [16] for this purpose. First, the creation of a dictionary of words in the data set was carried out and then it was further followed by the creation of a vector representation of the sentences according to the dictionary indexes. Initially, the Embedding layer was used to learn the embedding relationship between words in a sentence. Thereafter the output was passed to the LSTM layer and two Dense layers.

B. Feature extraction mechanism

When a new Intent was created by the user, he or she should define the required parameters. Mainly, for the feasibility of the study, eleven commonly used parameters such as name, URL, address, telephone no, date, count, color, location, email, time have been identified and defined them as data types with identical methods to identify them in given user input.

Since the Named Entity Recognition models for Sinhala [17] were not publicly available, a regular expression [18] based algorithms for feature extraction have been used here. The extraction method has been different for each datatype. A corpus of cities in Sri Lanka with a hash map-based search algorithm was used to identify cities in the given text. But to avoid searching each word in the text, a regex function was used to check for the most promising word. For example, it was identified that there is more chance for a word to be the name of a city that was following the words "ඉන්නෙ" or "ට" ("කොළඹ ඉන්නෙ"). Here two potential approaches have been identified and applied, as Eager Extraction Approach (EEA) and Lazy Extraction Approach (LEA). In EEA, the system was able to look for all possible features to extract in the given conversational text but in LEA, it looked only for the features which were defined and required for the identified Intent. Since the main disadvantage of EEA was higher computational power consumption, thus the study was continued with the LEA. The pseudo-code for LEA has been depicted in the Algorithm. 2.

Algorithm 1: RG pseudo code

```
Require: I //the intent to be evaluated
Require: text //user's message
Initialize features ← [], i ← 0

parameters ← I.getParameters()
foreach param ∈ parameters do
    /* Check the text for pre-defined
       patterns for parameter          */
    result ← parm.checkForParameter(text)
    if result   NULL then
        features[i] ← result
        i++
```

C. Response Generator (RG)

The first task of these modules was to check whether all the required parameter for the mapped intent was given by the user. If any parameter was missing, the RG module then asked it from the user. When all the parameters were set, RG may ask for a confirmation and then proceed with the intent as defined. For a given user message taken as an example, "මට සිකුරාද හවස 4.00ට වෙලාවක් වෙන්කරගන්න පුළුවන්ද?" assistant should be able to map it to the "Appointment" intent, extracted the details and parse a JSON object to RG. Response for the corresponding user message can be depicted as shown in Fig. 3. The pseudo-code for the RG has been depicted in the Algorithm. 1.

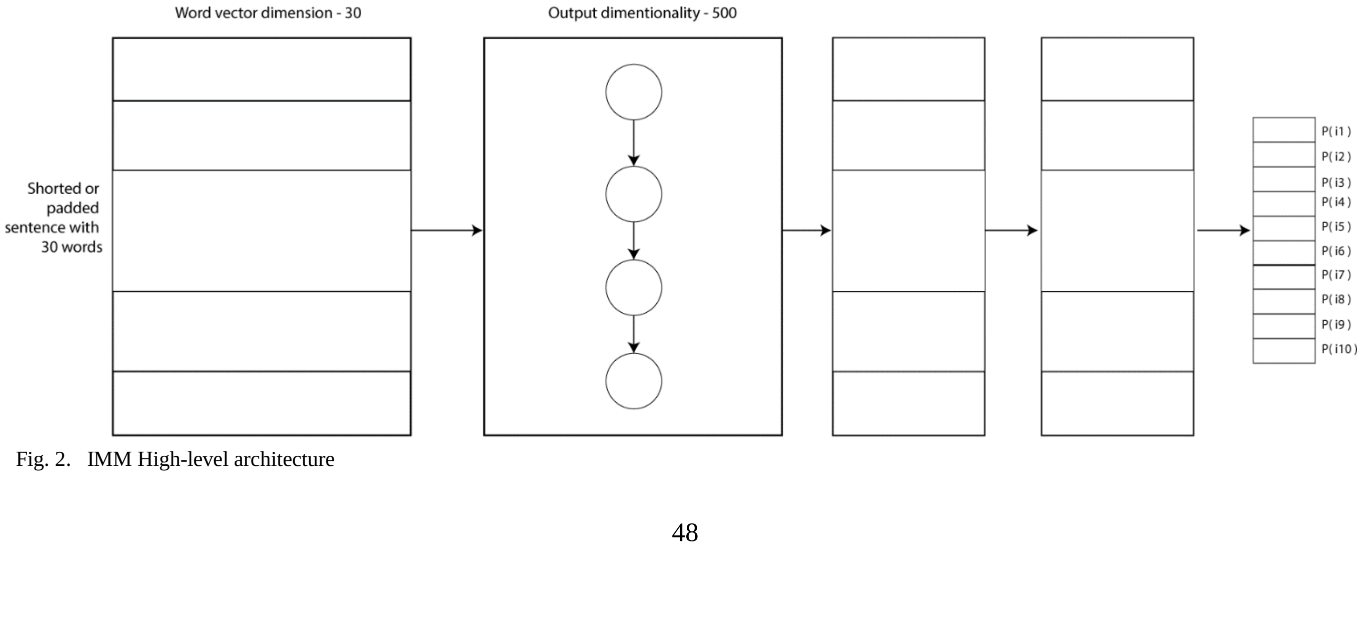

Fig. 2. IMM High-level architecture

```
{
    "intent": "Appointment",
    "details":{
        "date": "Friday",
        "time":"04.00pm"
}
```

Fig. 3. Sample user message with the response

Algorithm 2: LEA pseudo code

Require: I, the intent to be evaluated =0

Initialize $features \leftarrow [], i = 0$

if $I.isAllParametersSet == false$ **then**

  foreach $param \in I.getEmptyFeatures()$ **do**

    $f \leftarrow param.askFeatureQuestion()$

    $features[i] \leftarrow f$

    $i{+}{+}$

  /* After all parameters are set */

  if $I.getConfirmation() == true$ **then**

    $I.proceedIntent()$

  else

    $I.\ cancelIntent()$

D. Knowledge base

It is necessary to teach the computer program about the real world. The program should have knowledge about the business and its features to have a business-oriented and productive conversation with the customer. In this study, a database of information about the business was used as a knowledge base.

E. HTTP interface

There are some scenarios in which it is necessary to communicate with the business system. In 'cancelOrder' intent, the assistant's responsibility was to interpret the user message and then ask the checkout API of the business to cancel the order. The task has been accomplished by HTTP POST request to the checkout API of the business system.

F. Data acquisition

Since, gathering historical data was a difficult task, in this study data from social media were used for ease. Even with the slight differences in sentence patterns, it was figured that the embedding relationship between them was similar. Hence the consumer of this proposed framework might need to identify the related Intent for the scenario and the keywords bound to that. Thereafter, the dataset could be created by querying the keywords in social media platforms. For each social media platform, a unique data preprocessing method might be needed. Finally, all the preprocessed data should be presented in an excel sheet to be labeled by field experts.

V. APPLICATION

The proposed framework has been applied to make an assistant for the fashion industry. First, five Intents as 'cancelOrder', 'changeAddress', 'openingHours', 'askSizes', 'telephone' was separately identified and then the keywords and frequent forms of questions related to them were defined and categorized using survey and brainstorming.

Data collection has been carried out by using twitter standard API, Facebook Graph API and some were inserted manually without been biased. The collected dataset of 776 records, was then imported to Excel sheet which interfaces with experts from a variety of dialects, including graduates, teachers and other language experts. In preprocessing, first, all the emojis, URLs, hashtags, user mentions, and special characters were removed. Then, stop words were removed using a list of stop words [19]. After that, "mahaprana" characters were converted to their related "alpaprana" characters to simplify them. For example, the character "ඝ" was converted to "ග". In the final dataset,there are two columns as 'text' and 'class'. Data distribution of the dataset were shown in Fig. 4

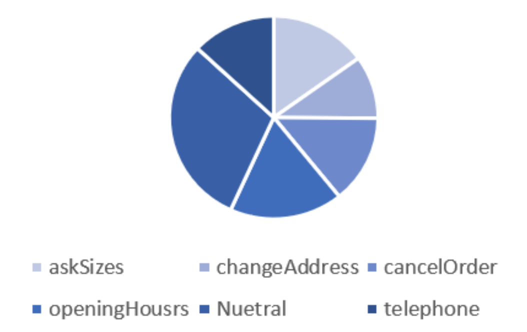

Fig. 4. Dataset distribution

A. Results and Discussion

To evaluate the model's performance, first, we evaluate the model with the test dataset of 112 records. Model's performances were shown using a confusion matrix (Fig. 5) with corresponding precision, recall and f1-scores for each class (Table. I). The overall model has shown an accuracy of 89.67. This model has shown higher recall value as expected because classifying them to the exact intent was critical. To ensure the practical usability of the overall application, a set of real users was asked to interact with the chatbot assistant. According to there feedback, the chatbot was able to come up with related responses with 75 percent of the time. We have observed that the model was performed well with short texts rather than longer ones and also it has performed well with linear conversations rather than non-linear conversations.

TABLE I. PRECISION, RECALL AND F1 SCORE

Intent	Precision	Recall	F1-score
askSized	0.90	0.82	0.86
cancelOrder	0.81	1.00	0.90
changeAddress	0.91	0.95	0.93
openingHours	1.00	0.82	0.90
telephone	0.96	0.96	0.96

VI. CONCLUSIONS

In this study, a framework for the contextual assistant, which can map user obligations to a predefined Intent has been proposed and implemented using the mentioned novel approach. The result of this study was promising the practicality of the approach. Main contributions from this study were:

- Recurrent Neural Network-based intent mapping algorithm
- The regular expression-based feature extraction mechanism
- A corpus of cities in Sri Lanka.

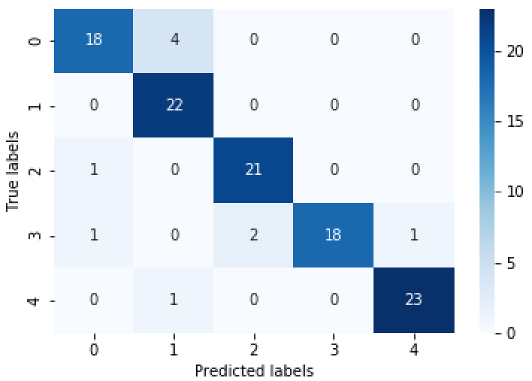

Fig. 5. Confusion matrix

This research study helped identify the key aspects of developing a contextual assistant framework especially for resource-poor Sinhala language and thus final output was a performing contextual assistant framework with a considerable accuracy rate to be used for the Sinhala language that could possibly be applicable for any business context.

So as future modifications it can be mentioned to follow up the ladder of maturity levels for the contextual assistant framework and target building more accurate maturity level incremented applications, especially for resource-poor languages, like Sinhala. The authors would like to suggest that the results could be further improved by using a word vector-based algorithm for the Sinhala language. Further improvements for this framework can upgrade usage of this framework to another variety of resource-poor languages.

REFERENCES

[1] Roar.media. [online] Available at: https://roar.media/english/tech/insights/the-rise-of-the-chatbot/ [Accessed 12 Nov. 2019].

[2] S. Janarthanam, P. Hall and V. Venigalla, “Hands-On chatbots and conversational UI development,” Birmingham: Packt Publishing, 2017.

[3] Evainmotion.com. (2019). EVA in motion. [online] Available at: https://evainmotion.com/ [Accessed 31 Dec. 2019].

[4] P. Jedrzejowicz, N. Nguyen, R. Howlett and L. Jain, “Agent and multi- agent systems: technologies and applications,” Berlin: Springer,2010.

[5] M. Mantha, Conversational AI: Design Build a Contex- tual Assistant - Part 1 — Sirius. [online] Sirius Computer So- lutions. Available at:https://www.siriuscom.com/2019/07/design-build-contextual-assistant/ [Accessed 29 Nov. 2019].

[6] Seylan.lk. Seylan Bank Expands Digital Customer Sup- port, Launches Seylan Facebook Messenger Chat Bot for the first time in Sri Lanka — Seylan Bank PLC. [online] Available at: https://www.seylan.lk/seylan-bank-expands-digital-customer-support-launches-seylan-facebook-messenger-chat-bot [Accessed 30 Nov. 2019].

[7] En.wikipedia.org. Sinhalese people. [online] Available at: https://en.wikipedia.org/wiki/Sinhalese people#cite note-Sinhala-1 [Ac- cessed 19 Oct. 2019].

[8] GeeksforGeeks. Introduction to Recurrent Neu- ral Network- GeeksforGeeks. [online] Available at https://www.geeksforgeeks. org/introduction-to-recurrent-neural- network/ [Accessed 15 Dec. 2019].

[9] B. Hettige, A. S. Karunananda, “First Sinhala chatbot in action,” *Proceedings of the 3rd Annual Sessions of Sri Lanka Association for Artificial Intelligence(SLAAI), University of Moratuwa*, 2006.

[10] B. Hettige, A. S. Karunananda, G. Rzevski, ”MaSMT: A multi-agent system development framework for English-Sinhala machine transla- tion,*” International Journal of Computational Linguistics and Natural Language Processing (IJCLNLP)*, Volume 2 Issue 7 July 2013.

[11] B. Hettige, A. S. Karunananda, “Octopus: A multi-agent chat-bot,” *Proceedings of 8th International Research Conference KDU* 2015.

[12] J. Weizenbaum, “ELIZA–A computer program for the study of Natural Language communication between man and machine,” *Communications of the ACM*, Vol. 9, No. 1, pp.36-45. 1965.

[13] ALICE, A.L.I.C.E AI Foundation, Available at: http://www.alicebot.org/, or http://alicebot.franz.com/. [Accessed 16 Dec. 2019].

[14] Assistant. (2019). Google Assistant, your own personal Google. [online] Available at: https://assistant.google.com/ [Accessed 16 Dec. 2019].

[15] Apple. (2019). Siri. [online] Available at: https://www.apple.com/siri/ [Accessed 19 Dec. 2019].

[16] Pitsilis, K. Georgios, Heri Ramampiaro, and Helge Langseth. “Effective hate-speech detection in Twitter data using Recurrent Neural Networks,” Applied Intelligence 48.12, 2018, 4730–4742. Crossref. Web.

[17] Dahanayaka, Jinadi & Weerasinghe, Ruvan., “Named Entity Recognition for Sinhala Language,” 10.1109/ICTER., 2014, 7083904.

[18] L. Ji, J. Guangyu, X. Aijun, and W. Yunzhen, “The automatic extraction of web information based on Regular Expression.,” *Journal of Software*, 12(4), pp.180-188, 2017

[19] Gallege, S. Analysis of Sinhala Using Natural Language Processing Techniques. [online] Citeseerx.ist.psu.edu. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.416.5173 [Ac- cessed 16 Aug. 2019].