A Study on a Chatbot Service Model Architecture using Open Source Chatbot Builders

Cheong Ghil Kim[*†]

[*†] Department of Computer Science, Namseoul University

ABSTRACT

Due to the development of IT technology and the on-going Coronavirus disease, non-face-to-face services have been activated. To overcome the inconvenience of non-face-to-face service, service providers have adopted chatbots as a way to feel like a human being. As the increasing chatbot services, chatbot builders have emerged, which can help non-developers to build them. Although its popularity has increased, its performance evaluation has not been conducted on such chatbot builders. In this paper, we implement a prototype chatbot that classifies hospital departments in the medical field using Dialogflow and Rasa, which are popular chatbot builders. By measuring the accuracy of the chatbot's classification of medical subjects, we evaluated the level of accuracy that the most used chatbot builder can have when they are used to build a chatbot service. The simulation results showed that Dialogflow had 87%, 65%, and 60%, and Rasa did 64%, 70%, and 63% in surgery dermatology, and otolaryngology, respectively.

Key Words : NLP, NLU, Chatbot, Chatbot builder, Rasa, Dialogflow

1. Introduction

As IT technology develops, it is changing in the form of combining with digital technologies such as AI, IOT, and Big Data in various fields. In addition, as the form of services provided in the face-to-face form has changed to non-face-to-face after the outbreak of the corona pandemic, the need for chatbots has increased. Compared to the existing face-to-face service, the non-face-to-face service has a possibility of receiving an uncomfortable environment from the user's point of view, and thus the user is likely to experience inconvenience. In order to overcome this problem, more and more companies are choosing chatbots as a medium for interaction with customers.

Due to various researches on chatbot technology, chatbots are providing services beyond just conversations, such as automated systems, consultations, and healthcare. However, in order to provide a chatbot service with high satisfaction, there are many performance differences depending on the accuracy of the NLU task mounted on the chatbot. Therefore, chatbot builders are often used to save development time and resources when developing chatbots [1, 2]. As mentioned earlier, the number of chatbot service using chatbot builders has increased, but the performance comparisons of them have not been conducted much. . In this paper, we implement a prototype chatbot that classifies hospital departments in the medical field using Dialogflow and Rasa, which are popular chatbot builders. By measuring the accuracy of the chatbot's classification of medical subjects, we evaluated the level of accuracy that the most used chatbot builder can have when they are used to build a chatbot service.

The structure of this paper is as follows. Chapter 2 describes the chatbot, the core of the chatbot service, and describes the chatbot builders Dialogflow and Rasa to be compared. Chapter 3 first describes comparison methods. Chapter 4 describes the simulation results and Section 5 concludes.

[†]E-mail: cgkim@nsu.ac.kr

2. Background

2.1 Chabot

A chatbot refers to artificial intelligence that can talk to people based on voice or text [3]. At the time of the first chatbot, Eliza [4], was possible with a simple interaction that responded only to certain words. Now chatbot is used for a variety of purposes, including cyber assistants who replace their role as a secretary, counseling chatbot for counseling, and chatbots that provide healthcare services to users.

Chatbots are largely classified into a retrieval-based model called a search model and a generative model called a generative model. It detects the intent of the user when they ask a question, searches for the best answer among the responses entered by the developer in advance. The retrieval-based model can only answer the entered query, but it can provide an accurate answer, and only the grammatically correct answer may be returned. Generative-model refers to a model that generates an answer that matches the user's query. It is an ideal chatbot in that it can answer questions not intended by the developer. However, it is not always possible to say that the generated response is syntactically correct and requires a lot of data for training.

2.2 Dialogflow

Dialogflow is a chatbot builder created by Google that helps to easily integrate a conversation interface into various systems such as mobile, web, and bots [5]. It is divided into ES used in small projects and flowchart-type CX used in large-scale chatbot projects. Even without knowledge, it is possible to implement a scenario with a GUI [5]. Due to the above characteristics, system development research using DialogFlow is being actively utilized in various fields.

2.3 Rasa

Rasa is a python open source chatbot builder software that aims to enable non-professionals to develop chatbot conversation systems. Rasa consists of Rasa NLU, which understands the user's query and performs intent analysis and entity extraction, and Rasa Core, which predicts according to the intent analyzed by NLU. Unlike Dialogflow, NLU and Core can be separated and trained. Fig. 1 shows how Rasa works.

The message entered into Rasa goes into the interpreter belonging to the Rasa NLU to understand the intent of the query and extract entities. Tracker is an object that saves the current conversation state. When the intent is grasped and entity extraction is completed in Rasa NLU, the policy is an object that selects the next action to be performed by Rasa Core. It can be seen that the correct answer is selected through this. If the action is selected through the policy, the current conversation state is saved again through the Tracker, and a message according to the action is sent to the user to inquire whether the correct answer has been derived. If an incorrect answer is derived, it goes back to the tracker and returns to the stage where the policy selects an action through the current conversation state [8].

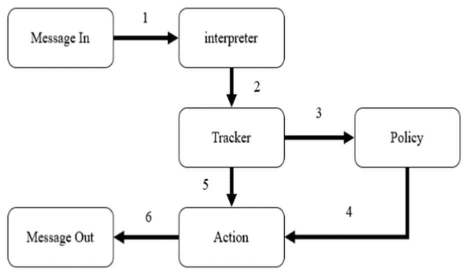

Fig. 1. Rasa Architecture.

3. Comparison of Dialogflow and Rasa

This paper compares the performance between Dialogflow and Rasa. Considering that the chatbot builder aims to enable non-developers to develop chatbot services, the construction difficulty is compared based on scalability, learning method, and environmental system construction constraints. The evaluation criteria are selected according to the criteria shown in Table 1.

Table 1. Performance comparison category

	Category	점수
Plat Environment	Cloud	3
	Local Computer	2
	Separate hardware required	1
Learning Method	GUI	3
	CLI	2
	Need to convert a separate learning module	1
Environmental construction constraints	No restrictions	3
	1 or 2 restrictions	2
	Restriction 3 or higher	1

The platform implementation environment was based on the degree of load on the computer. Cloud, which does not burden the service provider, was given the highest score, and the case where additional hardware was required was given the lowest score. The learning method was given a score according to the user difficulty of learning the model. The lowest level of difficulty was set to GUI, and the lowest score was set to the case of using a separate command such as programming. Finally, the constraint was given a score according to the number of environments that have dependencies when building the environment. Dialogflow and Rasa, which are machine learning-based chatbot builders, are used for performance comparision. As a learning data, about 1,300 data were composed by crawling the medical questions of Naver intellectuals, and the data set removed the data that was determined to be related to special characters, gaps, and classification and separated by 1,200 learning sets and 100 verification sets. Learning and verification were conducted.

4. Simulation

In this paper, a qualitative and quantitative comparison of two chatbot builder platforms, Dialogflow and Rasa, was performed. Qualitative measurement was classified into three categories: platform, learning method, and environment construction constraints.

Table 2. Qualitative Comparison

Chatbot builder	Category	Classification	Score
Dialogflow	Plat Environment	Cloud	3
	Learning Method	GUI environment	3
	Environmental construction constraints	None	3
Rasa	Plat Environment	Local	2
	Learning Method	Separate instruction + CLI	1
	Environmental construction constraints	Python 3.8 and below, Rasa, Spacy,Tensorflow, etc	1

Since Dialogflow can build an environment only with a Google account in the Cloud environment, the score is given as above. In addition, we give 3 points because learning is conducted on a GUI-based web page. For Rasa, it is 2 because it's a chatbot builder platform is on local computer. In the learning method, since training data must be input in Rasa command and Yml file, separate learning is required, so point 1 is given. In the environment construction, there is a restriction on the Python version, and not only the command changes according to Rasa version, but also the constraint according to the version of Spacy is given point 1 to be observed. In the case of environment construction, in the case of Rasa, we give point 1 because there are more than 3 version restrictions for python and libraries with dependencies. For quantitative measurement, 1,200 training sets and 100 datasets are used for each classification department. Dialogflow is implemented through GUI, and in the case of Rasa, it is implemented on Windows 10, Python 3.8.2, Rasa 3.0, and Spacy 3.0. Table 2 shows the result of measuring the accuracy using two chatbot builders.

Table 3. Chabot builder accuracy

Chabot builder	Classification	Accuracy
Dialogflow	Surgery	87%
	Dermatology	65%
	Otorhinolaryngology	60%
Rasa	Surgery	64%
	Dermatology	70%
	Otorhinolaryngology	63%

5. Conclusion

As the increasing chatbot services, chatbot builders have emerged, which can help non-developers to build them. In this paper, to measure the qualitative and quantitative performance that can be obtained when implementing a chatbot service using chatbot builders, we implemented a chatbot that categorizes the three departments of surgery, otolaryngology, and dermatology with Dialogflow and Rasa, which are representative chatbot builders. The classification accuracy of the chatbot was measured. As a result of measuring the accuracy, Dialogflow had 87%, 65%, and 60% in surgery, dermatology, otolaryngology, respectively. This result shows that DialogFlow has a higher accuracy measurement than Rasa.

Acknowledgement

Funding for this paper was provided by Namseoul University year 2021.

References

1. A. Abdellatif, K. Badran, D. Costa, E. Shihab, A Comparison of Natural Language Understanding Platforms for Chatbots in Software Engineering, IEEE Transactions on Software Engineering , 2021, pp 1-1.
2. HaeKyung Chung and Jungjo Na, Development of Wine Recommendation App Using Artificial Intelligence-Based Chatbot Service, Journal of the Semiconductor & Display Technology, Vol. 18, No. 3. September 2019, pp. 93-99.
3. T. Lalwani, S. Bhalotia, A. Pal, S. Bisen, V. Rathod, "Implementation of a Chatbot System using AI and NLP" IJIRCST (International Journal of Innovative Research in Computer Science & Technology, Vol. 6, 2018.
4. J. Weizenbaum, ELIZA A Computer Program For the Study of Natural Language Communication Between Man and Machine, Communication of the ACM, 1966, Vol. 9, pp. 36-45.
5. A. F. Muhammad, D. Susanto, A. Alimudin, F. Adila, M. H. Assidiqi, S. Nabhan, Developing English Conversation Chatbot Using Dialogflow, 2020 International Electronics Symposium (IES), 2020.
6. Y. B. Yoon, "Prospects of Using AI Chatbots in Tech Speaking in Primary English: With Special Reference to Dialogflow" The Journal of Korea Elementary Education, 2021, Vol. 32, pp. 15-28.
7. S. Salvi, G. V, S. K. S, Jamura: A Conversational Smart Home Assistant Built on Telegram and Google Dialogflow, 2019 Region 10 Conference IEEE, 2019
8. T. Bocklisch, J. Faulkner, N. Pawlowski, A. Nichol, "Rasa: Open Source Language Understanding and Dialogue Management" 2017, arXiv:1712.05181.

접수일: 2022년 9월 21일, 심사일: 2022년 12월 6일,
게재확정일: 2022년 12월 12일