

Primary Study for dialogue based on Ordering Chatbot

Ji-Ho Kim¹, JongWon Park¹, Ji-Bum Moon¹, Yulim Lee¹, Andy Kyung-yong Yoon^{2,*}

Abstract

Today is the era of artificial intelligence. With the development of artificial intelligence, machines have begun to impersonate various human characteristics today. Chatbot is one instance of this interactive artificial intelligence. Chatbot is a computer program that enables to conduct natural conversations with people. As mentioned above, Chatbot conducted conversations in text, but Chatbot, in this study evolves to perform commands based on speech-recognition. In order for Chatbot to perfectly emulate a human dialogue, it is necessary to analyze the sentence correctly and extract appropriate response. To accomplish this, the sentence is classified into three types: objects, actions, and preferences. This study shows how objects is analyzed and processed, and also demonstrates the possibility of evolving from an elementary model to an advanced intelligent system. By this study, it will be evaluated that speech-recognition based Chatbot have improved order-processing time efficiency compared to text based Chatbot. Once this study is done, speech-recognition based Chatbot have the potential to automate customer service and reduce human effort.

Key Words: Chatbot, Artificial Intelligence, Speech recognition, Sentence analyze, STT, TTS, dialogue emulation.

I. INTRODUCTION

In recent years, AI has been the mainstream technology of the 4th Industrial Revolution and has attracted the most attention in businesses, government policy and IT industry. However, it seems difficult for SMEs, Small-Giants and start-ups to follow the precise trends and indicators of the era in the 4th Industrial Revolution [2]. Among the AI industry groups, deep-running and speech-recognition are the most popular areas. Many companies invest and develop in this field, but it is not easy for start-ups and SMEs to survive. Many global companies have successfully developed interactive dialogue-based Chatbots using speech-recognition, but Korean speech-recognition based Chatbots are still poorly recognized than English Chatbots. The reason is that the UI was given more weight. The other is that the research perspective was different from the global one, due to AI research in Korea started later than global, and focused on only commercial aspects [1-3].

The goal of this study is to acquire user-oriented AI technology that has user-friendly interface and user-oriented usability. Therefore, it is necessary to implement the Chatbot system that can lead technological innovation by using accurate sentence analysis, syllable analysis, and automatic estimation word prediction method based on

speech and text, and to secure technological ability to operate easily in Korean [2].

II. PREPARATION

2.1 Context of speech processing

Digital speech processing is classified into STT and TTS, that is, recognition and synthesis. Therefore, STT converts speech to text, and vice versa is TTS. STT is one of the hottest areas that have been continuously studied by many researchers. It also serves as an intermediary for dialogue between humans and robots. The reason why STT is attracting attention in AI industry is that it is applicable to various fields and it is one of the good tools to increase profit by fast and clear data processing [3,8].

In order to achieve the goal of this study, mobile applications developed from outside have been analyzed and the feasibility of utilizing the API has been examined. Based on this, practical problems were analyzed and the direction of development was determined. The external APIs to be analyzed is developed by best IT player in Korea such as KaKao, Naver, and ETRI. The advantage of the KaKao STT was obvious speech-recognition and clear and sophisticated word processing. However, it is not easy to systematic techniques in this study by benchmarking [2].

Manuscript received September 19, 2018; Accepted September 26, 2018. (ID No. JMIS-2018-0047)

Corresponding Author (*): Andy Kyung-yong Yoon, xperado@usmp.pe

¹YAP Company, Seoul Korea, jiho.kim@yap.net, jongwon.park@yap.net, jibum.Moon@yap.net, yulim.lee@yap.net

²Special School of University of San Martin, Peru, xperado@usmp.pe

The advantage of ETRI STT was that it was able to analyze Korean sentences obviously and speed up data processing, but it was difficult to transplant to this study compared to other STTs because it used the basic interface.

For this study, various mobile devices such as Galaxy S7, Galaxy Note8, LG G6, Galaxy Tab A and LG V10 have been prepared. The main items tested were speech recognition rate and processing speed for each model. Further, WiFi and three mobile carrier's networks were used to understand the diversity of the network [7].

MySQL and LINUX (Ubuntu, SSD Samsung 850, x86_64bit, intel-core i5-6600 @ CPU 3.30GHz) were used as database and server. The target application is a beverage ordering system, and the Chatbot engine will be implemented to use this application. The emphasis was on developing an engine that would make it easier to upgrade the application in the future. The reason for using many devices and servers during development is to minimize the inherent characteristic problems of each device [1].

2.2 Basic Chatbot model

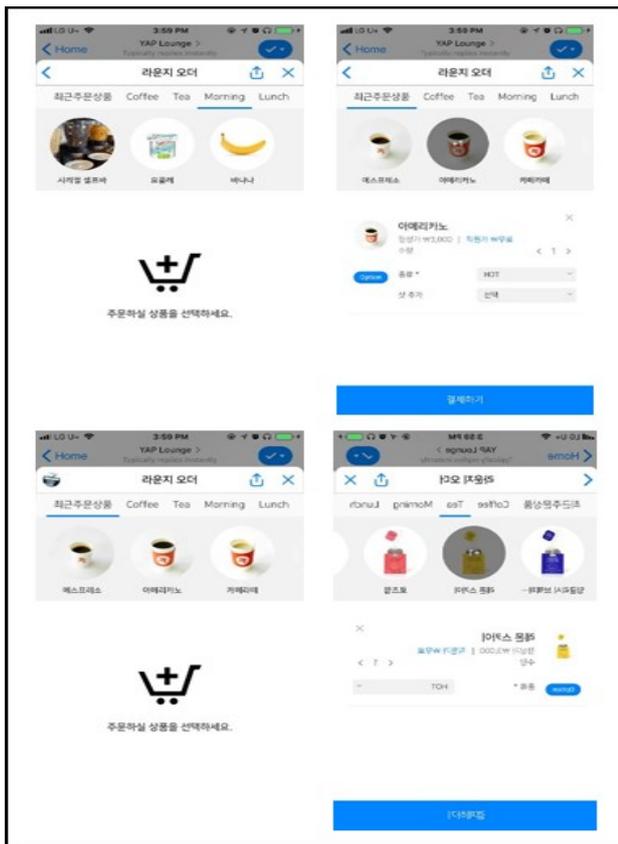


Fig 1. WebChatBot

The starting point of this study is the WebChatBot as shown in Fig. 1. This is one of the basic Chatbot systems. Hereinafter this shall be referred as a menu-driven method for convenience. The advantage of this system is that it gives users a clear choice. It is also an important advantage that users are encouraged to select only the choices shown

in well-organized stories. However, the administrator had to make all standard choices in advance, which caused frequent network communications and took an average of 1~2 minutes to get the final results. Thus, a persistent problems are a slow response and consuming lots of network resources [3-5].

In order to overcome the problems of the menu-driven method, the study for speech-recognition method system has begun. If "Americano" is input by speech, it is designed to be able to distinguish objects, actions, and preferences as shown in Fig. 2.

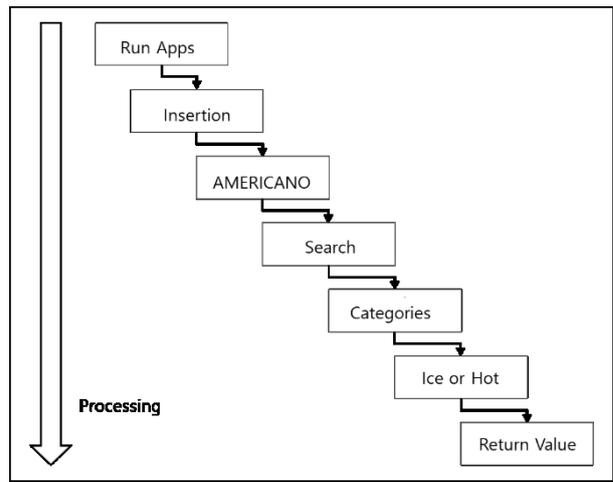


Fig 2. Basic Process concepts for Application

In other words, the menu-driven method consumes a long time and a lot of network resources because it is a method of sequentially generating an answer. On the other hand, the speech recognition method is designed to differentiate the process within one second as soon as the speech is input, and to receive accurate results even with a small number of searches [6].

III. DESIGN

3.1 Service Flow

As shown in Fig. 3, the process is expanded after the user clicks on the application. The user service flowchart to show the main screen at launching the application so that the conversation can be made by voice. In addition, the user can select the food in the menu section by clicking on Text, not speech only. Application aimed for text and speech. [5]

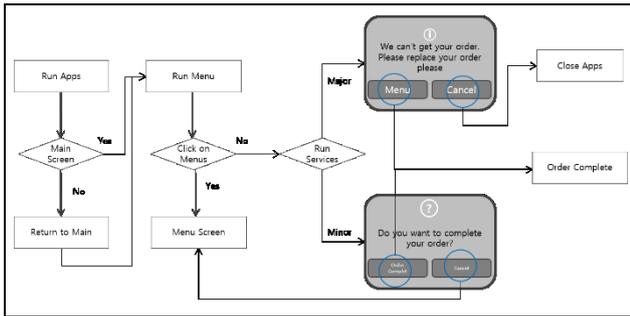


Fig 3. Flow chart, flow shows entire application in main screen.

3.2 Data flow diagram

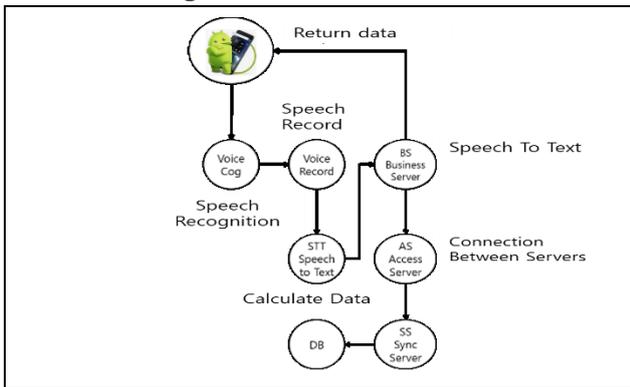


Fig 4. Data Flow Diagram, Diagram designated as UX's first-step to final -step.

It is an extension of the overall module drawing of data in to the Fig. 4. Servers can be combined as Synchronized Server, Access Server, and Business Server etc. Each server has different tasks handling. The unique role of Business Server can be explained as a “collection of algorithms”. It is also a conversation between the user and the server. The Access Server gives connection from the Business server and a link from the Synchronized Server. Synchronized server is one of the most important servers. Synchronized server has 2 basic roles in logic. First is connecting to databases. Second, they collect atypical data from Access server and return calculated values to Access Server. In this study, application allocated more than one server because of avoiding bottleneck in the network. And also during development, operator can monitor dataflow from servers [3-4].

IV. COMPOSITION

4.1 Syllabic Analysis Method

In entire process, the blank exception handling of the received value will preferentially. Handling can be supported by program was used. And also system will return value as no spaces from insertion. For example,

When “아~이스” (In English, Ice) entered into the system, system will recognize as “아이스” with no spaces. The system has main constraint from insertion. The entire insertion of synonyms, system must collect at least two cases from insertion. When “커피” (In English, Coffee) entered into the system, the system will represent entire categories related with coffee. And also system will be returned as “Smoothies”, “Lattés”, “Frappuccino, etc.” so user can perceive their purposes. To do synonyms method, Algorithms classified into 4 cases “A.M.E.R.I.C.A.N.O” into 4 cases. “A.M.E”, “A.M.E.R.I”, “A.M.E.R.I.C.A”, “A.M.E.R.I.C.A.N.O”. When system continued to erase 4 cases, system will extract until “C.A.N.O”. Because, that is irregular data from user's insertion. In this research, the system highly recommend to avoiding irregular data. It depends on Server's accurate result value can be derived based on high perception. To do further algorithms on other methods, system will try to find patterns from data. The patterned data always saved in Mass Map as shown in Fig. 6. The main reason for finding data will be “collect information of the Object (target).”

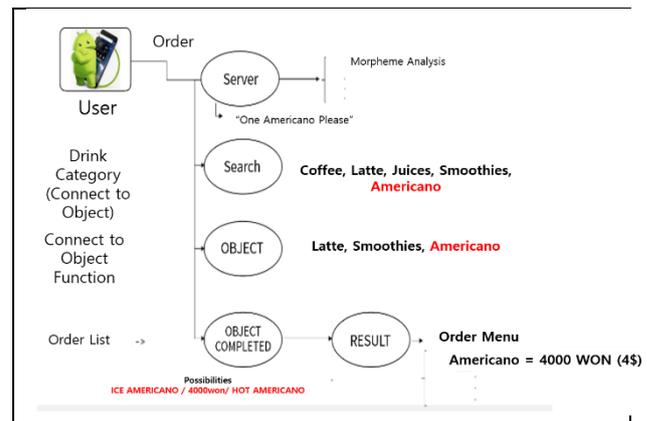


Fig 5. App user insert OBJECT to system and system search the right object.

4.2. Acquisition of user input prediction data and mass MAP generation technique

In this method, System will create one of a big map to collect ASCII code from mass map method. Mass map method calculate data structure algorithm that retrieves the basis of deliverables of the input prediction data. Retrieval method using a utilization scheme that extracts values from *n* parts of data around an infinite loop as a measure to clear data. In this method, system will distribute by one letter from insertion.

And also System will collect calculated data used by mass map. Based on algorithm, Data structure creates a multi-dimensional array to put ASCII code. For example, “C.O.F.F.E.E” can be specified C = 99, O=111, F = 102, F = 102, E = 101, E = 101, the total mass can be 616 in ASCII value. As we calculated those functions, we can specify

patterned data. "Input Mass / increment = index of mass Map." As shown in Fig. 6 and 7, System created multi-dimensional arrays to extract confidential patterns from data and system can access or find data in less search.

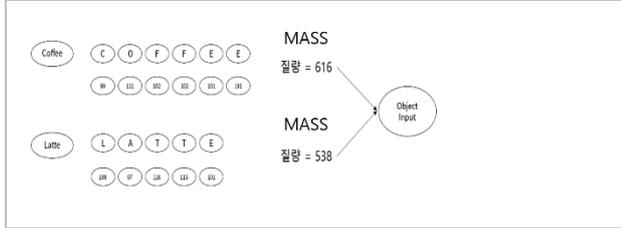


Fig 6. Each Object (target) contained unique Mass. Figured mass order by ASCII CODE.

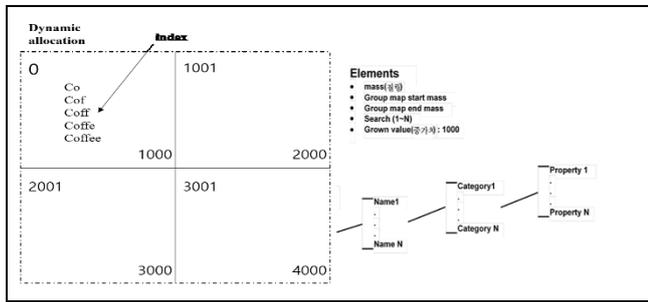


Fig 7. Each Mass located in specific MAP.

V. EVALUATION AND DISCUSSION

In this study, we performed the task of deriving the result with only 'Object' as the pilot version. Thus, we evaluated how many objects are in a sentence, and the ability to recognize and sort all of these objects. As mentioned earlier, the final goal was to classify this sentence into three if the sentence was entered. The first classification is an object. Therefore, it is difficult to interpret the whole sentence which is entered by speech only on this study. Therefore, in this basic study, it is only to find all the objects, recognize them, and evaluate the rate of recognition.

For example, when the speech "아이스 아메리카노 주세요" (Can I have an Ice-American?) is entered, the main object is to find a "Coffee". If "Americano" is found, this is a 100% probability, due to which will be replaced by coffee. Of course, this sentence is incomplete. The complete sentence would be "아이스 아메리카노 커피 1 잔 주세요" (Can I have a cup of Ice-America Coffee?). Or rather "~~카노 주세요" (~~ kano please) is somewhat unclear, but if it will be deduced the word "kano" as "Americano" and also replace it with "Coffee" and recognize it as an object, The probability is 100%.

In natural languages, objects always follow a lot of modifiers. For example, 'dark', 'cool', 'warm', 'delicious', or 'not too much bitter'. This modifier thus changes the language to various feelings such as smooth, rough, dry, and so on. However, this modifier is a very cumbersome noise for algorithms. Therefore, filtering out all these modifiers is one of the important algorithms. Assuming that the just a dry sentence "커피 한잔 주세요" (Please give me a

cup of coffee) is entered, this sentence is classified as "Object", "Action", "Preference", ie, "coffee", "give", "a cup". At this time, if only "coffee" is recognized as an object, the probability is 100%, but if "a cup" is also recognized as an object, its probability of recognition falls to 50%.

Therefore, although "cup" is recognized, algorithms must filter them out as "Preferences" rather than "Objects". To evaluate this, 25 natural language sentences were enter as shown in Table 1, and the recognition rate of the object was evaluated.

As shown in Fig. 8, the dash line represents the total number of objects, and the dot line represents the number of recognized objects. And the solid line represents the probability.

Table 1. Experiment Scripts

No.	Scripts	Total(n) Object	Recognized	Failure	Probabilities
1	아이스 아메리카노 한잔 주세요 (Give me one ice Americano Please)	1	1	0%	100%
2	따뜻한 아메리카노 한잔 주세요 (Give me hot Americano)	1	1	0%	100%
3	아메리카노 시럽 추가해주세요 (Please give me Americano, and add syrup)	1	2	50%	50%
4	아메리카노 한잔이랑 물좀 주실래요 ? (Give me one Americano with glass of water)	2	2	0%	100%
5	핫초코하나주세요 (Give me Hot Chocolate)	3	2	25%	75%
6	아이스 아메리카노 미지근하게해주세요 (Give me Americano, and less ice)	4	2	50%	50%
7	따뜻한 아메리카노 헤이즐넛 이랑 물이요 (Give me Americano with Hazelnut and glass of water)	4	4	0%	100%
8	그냥 커피 주실래요? (Just give me one cup of coffee)	1	1	0%	100%
9	카노 투 샷 추가요 (cano, add 2 shots)	3	1	75%	25%
10	리카노 (Give me licano)	1	1	0%	100%
11	노 한잔 주세요 (Give me one no)	1	0	100%	0%
12	아무 아메 주세요 (Give me one Ame)	2	1	50%	50%
13	메아 주세요 (Give me mea)	1	1	0%	100%
14	리아노 주세요 (Give me Liano)	1	1	0%	100%
15	나한테 커피좀래? (Give me coffee)	1	1	0%	100%
16	주스 한잔 좀래? (Give me juices)	1	1	0%	100%
17	초콜릿 우유 한잔만 줘 따뜻하게 (Give me hot-chocolate milk)	4	3	25%	75%
18	달달한 커피 좀래? (Give me something sweetened coffee)	2	1	50%	50%
19	달달한 아메리카노 있어요? (Is there any something relate with Americano)	2	2	50%	50%
20	아노 한잔 주세요 (Give me ano)	1	1	0%	100%
21	달달 주세요 (Give me something sweetened)	0	1	100%	0%
22	약간 쓴 커피 주실래요? (Give me something bitter)	1	2	50%	50%
23	추천 해주세요 (Please recommend something)	0	0	100%	0%
24	혹시 콜라 있어요? (Do you have Coke)	1	1	0%	100%
25	메노 있나요 ? (Do you have any Meno)	1	1	0%	100%

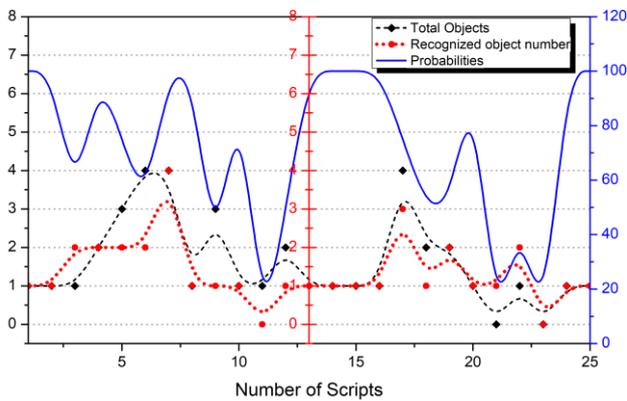


Fig. 8. This figure contains specific insertion, category, object, output and its order by probabilities

As described hereinabove, the recognition rate is close to 90%. Of course, this can be a high recognition rate because it is only for objects. However, it is expected that "Action" and "Preference" will show high recognition rate.

VI. CONCLUSION

In this study, the Korean sentences are defined in exactly three fields as Action, Object, and Properties. The result of object derivation performed with n numbers of algorithms but the system could not interpret it as a complete AI form. And also calculated-value could be little bit vague. However, if study added "Action" and "Properties", system can make dialogue with fully-constructed paragraph.

This study is only a part of AI application. Due to interactions between human and AI, Our future study aimed to build perfect chatbot application for the ordering system. In the future study, we would expect that users can be able to use our technology and skills to have conversation with AI.

REFERENCES

- [1] DongA Park. "A Study on conversational Public Administration Service of the Chatbot Based on Artificial Intelligence" *Journal of Korea Multimedia Society* Vol. 20. No. 8, August 2017(pp. 1347-1356).
- [2] Sumin Choi, Yongsoon Choi. "Analysis on the Conversational commerce Service Interface of the AI Chat-Bot Based on Mobile Messenger Apps" *PROCEEDINGS OF HCI KOREA 2017* No. 2, February 2017(pp. 237-240).
- [3] Eric Atwell, Bayan Abu Shawar. "Using dialogue corpora to train a chatbot" Conference: *Proceedings of the Corpus Linguistics 2003*

conference (pp. 681-690), At Lancaster University

- [4] A.F. van Woudenberg, "A Chatbot Dialogue Manager", M.S. theses, Open University of the Netherlands Faculty of Management, Science and Technology, June 17, 2014.
- [5] Jack Cahn, "CHATBOT: Architecture, Design, & Development", University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science April 26, 2017.
- [6] Bayu Adhi Tama and Kyung-Hyune Rhee1, *A Comparative Study of Phishing Websites Classification Based on Classifier Ensembles*. *Journal of Multimedia Information System* VOL. 5, NO. 2, June 2018 (pp. 99-104).
- [7] Tae Hun Hwang1, Jin Heon Kim, *An Approach to Improve the Contrast of Multi Scale Fusion Methods*. *Journal of Multimedia Information System* VOL. 5, NO. 2, June 2018 (pp. 87-90).
- [8] Sadaoki Furui, *Digital Speech Processing, Synthesis, and Recognition*, Tokyo Institute of Technology Tokyo, Japan.

Authors



Jiho Kim has received his BS degree from Indiana University-Bloomington, IN, U.S. Currently, he is the scientist in R&D Centre at YAP Company, Korea. His main research areas are geofencing engine, in-door positioning system, and speech recognition based chatbot engine. Further, his interests are AI related application and terminals.



JongWon Park has received his BS degree from SungSil University, Korea, in 1994. From April 2012 to June 2016, he joined in the Giga-geofencing service engine (KT) and O2O engine service research team at Immersive co. He was the Principal Research member in Immersive Co., Ltd., Korea. His main research area is the real-time server processing algorithms for geofencing engine. Currently, he is the Chief Scientist in R&D Centre at YAP Company, Korea.



Yulim Lee is involved R&D Centre at YAP Company, Korea after receiving of his BS degree. Currently, her main research areas are geofencing engine, indoor positioning system, and speech recognition based chatbot engine. Further, her interests are AI related application and game engines.



Jibum Moon is involved R&D Centre at YAP Company, Korea after receiving of his BS degree. His current research areas are dialogue based ordering chatbot engine, BLE radiomap based indoor positioning system, and geofencing system. He interests AI related big-data processing application, meta-data processing and server programs.



Andy Kyung-yong Yoon completed his MS and Ph.D. degree from the Yonsei University, Korea. He is currently a Professor in the Professional school of Electronic Engineering at San Martin University, Peru, also the CTO of YAP Company, Korea where his main activities include research, undergraduate and postgraduate training. His research interests include Mobile Agent Systems and AI related O2O application, and Indoor Positioning System.