

Intelligent Chatbot for guided Navigation of a Repository Content: The State of Art Survey

Anjali R. Mishra¹, Shreya R. Nahar², Shruti S. Shinde³, Shruti S. Sapre⁴

^{1,2,3,4}Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India

Abstract—Technology brought revolutionary changes in our lives. There have been many learning styles and knowledge improvement methods. A chatbot is chat robot which is a computer program and conducts conversations between human and machine which offers numerous services via chatting. Here we design a system which answers to questions from a repository of a content intelligently using Artificial Intelligence Markup Language (AIML) and Natural Language Processing (NLP). The users get answers to their questions about information in the book-given as input to the system, even without reading the book. The chatbot is the interface which will send message and get canned response and the conversation is stored in Firebase. Using a mind mapping tool we create and present models which interprets the questions and maps them to the model. To have humanlike chat with a bot Artificial Intelligence is essential. AIML is an XML based markup language meant to create artificial intelligent applications.

Index Terms—Artificial Intelligence Markup Language (AIML), Natural Language Processing (NLP), Chatbot, Pattern matching.

I. INTRODUCTION

Intelligent technology advancements in today's era are not only "an extension to human hand", but also "an extension to human brain". One of the important goals of Human Computer Interaction is to make computer more proactive, more hard-working and doing more work. So, human can do jobs as few as possible, more easily and conveniently. A harmonious HCI should make computer more intelligent so that user can do less work to achieve functions. Intelligence technology makes interactive ways faster, more natural and more convenient [1].

The demand of web services is increasing day by day, where everything is now getting associated with web. It is very convenient to avail everything at the user's doorsteps. There is a huge variety for customer services available like chat support services, phone services.

In the computer age, there are a vast quantity of documents rapidly created to meet user's needs. There are large amount of digital documents on the Internet, which are used in various different ways. All the documents containing the information support many type of decision-making. However, many questions are still unanswered, which brings the idea of constructing a system for questioning and answering doubts. This clearly brings us to see that the obstacle of building such system which is a knowledge provider [6].

Nevertheless, in several condition [7] e.g. in vacation or during a rest, the seller may be unable to communicate with the customer. But for a user, it may take time to answer customer's doubts. As the number of consumers increase the waiting time increases as well, which results in poor customer satisfaction. As a solution, the seller can employ several customer services to answer any questions from customer. However, it involves additional costs for paying customer service officers. As an alternative and better solution, we can develop a system that can receive and understand the question from customer and automatically, give the best response with respect to that question.

Automated live chat support service is the solution for such problem. Chatbot will take care of user queries and problems. It will give the solution for the same. A chatbot is designed to answer both general questions and FAQs about a particular domain [7].

Fortunately, in a last decade, many researchers are actively involved in the area of Artificial Intelligence and its related research topic including machine learning and natural language processing (NLP). With those advancement, we can develop a chat robot, or chatbot which is an automated program that can interact with its communicating counterpart without any human intervention. In order to build such a program, there are several challenges arise including : how to understand the asked question even though it contains any mistyped word, and then, select the best among several possible responses[8].

Building an intelligent Chabot which can adept to human brain like capabilities of memorizing, accepting user's preferences and providing relevant response.

II. CLASSIFICATION

Chatbots are the programs that are designed to simulate conversations with humans and respond to different input modulities such as natural language text in human like manner. Its advantage is the possibility of performing repetitive tasks given specific commands or requests while it gives the impression that it can understand and give answers Chatbots can be classified into two types according to the kind of model that can be used to produce its answers.

1. Deterministic model-those that provide a response based on a fixed set of possible answers.
2. Generative model-those that attempt to generate responses

by keeping posterior information over all possible answers [2].

In the first category, we can find systems that use pattern-matching techniques which are used to select the closest rule to the current user's turn and then pick an answer from the pre-defined set of possible options. Good examples of this kind of chatbots are the psychotherapist Eliza created in 1969 by Joseph Weizenbaum [3], which used a set of only 53 pattern matching rules, or A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) developed by Richard S. Wallace, [4] that won the Loebner's prize in 2000 and 2001 by using more than 50K rules. A.L.I.C.E. also known as Alicebot, is a natural language processing chatterbot: a program that manages conversation with a human by applying some pattern matching techniques to the user's input.

Advantages:

- Accessible anytime, anywhere.
- Handling capacity.
- Customer satisfaction.
- Cost efficient.
- Flexible.

Limitations:

- They can only provide answers that are defined in its heuristics.
- The answers may be inappropriate to particular inputs given their simplicity on handling the dialog context.
- They can be difficult to maintain as the number of rules increase.

In the second category, the chatbot does not make use of hard coded question/response pair but instead it completely produces a response based on a generative model trained using machine learning algorithms. These agents are capable of adapting themselves to the dialog context without the need for the developer to anticipate and provide an answer for all possible user inputs. Such kind of chatbots uses a very small set of system states and actions allowing the usage of reinforcement learning techniques. More recent approaches based on neural networks like generate responses selecting words from probability distributions by using methods like beam-search and encoder-decoder networks which are trained to simulate the behavior of the human sequence of turns [2].

Advantages:

- Improved accuracy.
- Gaining insights.
- Better understanding.
- Response based
- Keeps posterior information.

Few applications of chatbot are enlisted below:

1. Health based chatbots-Two most popular AI-based health chatbots are Dr. A.I. by HealthTap and Melody

by Baidu which automate work of Doctors.

2. Social networking based chatbots-Swelly is a chatbot which is used in facebook messenger.
3. Technology based chatbots-HP Print Bot effortlessly prints documents.

TABLE I
 DIFFERENCE BETWEEN DETERMINISTIC AND GENERATIVE MODEL

S. No.	Characteristic	Deterministic Model	Generative Model
1.	Response	Based on fixed set of possible answers	Generated by keeping previous data
2.	Posterior information	Does not store data	Stores data
3.	Accuracy	Less accurate	More accurate
4.	Understanding	Comparatively less understanding	Better understanding of user's questions
5.	Maintenance	Difficult	Easy
6.	Handling capacity	To less extent	To greater extent
7.	Examples	ALICE, ELIZA	Dr.A.I. by HealthTap



Fig. 1. HP print bot



Fig. 2. Luigi-Fiat chatbot

4. Travel based chatbots- Ticketbot for Airasia helps you to find prices for any Airasia route using simple language commands.
5. Banking based chatbots- Ask Me, Citibank India is built and designed to help users with their queries.Andrew HSBC

handles customer service work HSBC's commercial clients in UK.

6. Food based chatbots-DOM the Pizza Bot helps users in ordering Domino's Pizza from their facebook messenger account.
7. WiFi based chatbots- Hello Barbie, MATTEL is a companion chatbot built into wifi enabled talking barbie doll and has the ability to remember things from previous conversations and keep users engaged.
8. Automotive based chatbots- Luigi, Fiat Argentina is designed to help new Fiat owners after they have already made their purchase and also buyers make a decision.

III. METHODS

A. Latent Symantic Analysis (LSA)

We have two types of answers like pattern based answers by AIML and semantic based answers by LSA. The system will automatically route to LSA block if AIML is not able to give answer for the particular user input. User can ask any questions. We used the template data of ALICE to run our AIML model [7].

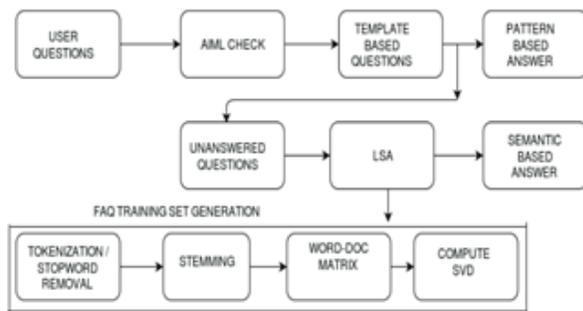


Fig. 3. Latent symantic analysis

Obtained Results: The chatbot is designed in such a way that for single template, it gives random responses. LSA based questions are giving correct responses. In future, the chatbot will be able to answer more general questions by defining more templates and patterns.

B. Naive Bayesian Classifier

The chat agent accepts user's query and extracts the keywords i.e. nouns and verbs from the question using a lexical parser. Then these keywords are compared with the category list which is present in database. The Bayesian probabilities are obtained for all categories in the list. This is done for reducing the search space so that the comparison with the actual question database is done easily. With the category selected, the keywords are then compared with the questions under the category using Bayesian probability theory. The answer to the question which has the highest posterior probability is then fed into the text to speech conversion module and thus the student receives the answer to his question as a voice response [9].

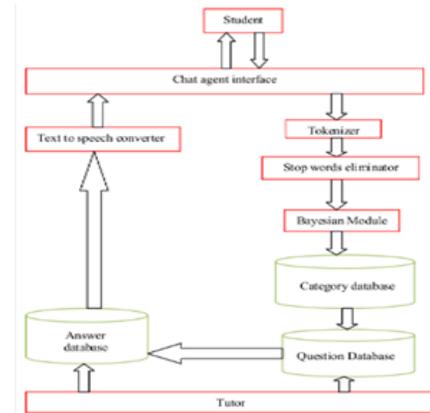


Fig. 4. Naive bayesian classifier

Obtained Results: The result here is obtained using 3 types: Using a single category, the system was tested using a self-made knowledge base having a single category and it was found that the system was capable of identifying the correct match most of the time. Using multiple categories, the system was then enhanced with knowledge on three categories and tested. Self-categorization: The system is implemented in such a way that it applies the Bayesian algorithm on the data that is given by the tutor and categorizes it by itself [9].

C. Optical Character Recognition (OCR)

It is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine-encoded text. Scanning and comparison technique are required to identify printed text and numerical data. It avoids the need to retype already printed material for entry of data. OCR software attempts to identify characters by comparing shapes to those stored in the software library. The software tries to identify words using character proximity and will try to reconstruct the original page layout.

Obtained Results: Chatbot is known as an assistance which provides information without need of human. This web-based Chatbot system enables user using in many factors: in Call center service, it can be used to answer the frequently ask questions. In education field, it can be used as a teacher, and many areas, it can provide appropriate guide assistance to make the right decision [6].

D. Question Generation via over generating Transformations and Ranking

It is a framework for generate a rank set of fact-based questions about the text of a given article by first, this approach compose of three stage process:

- Stage-1: Extract and derive declarative sentences from a source text;
- Stage-2: Transduce declarative sentences into questions using declarative, general-purpose rules;
- Stage-3. Statistically rank the output of over generating stages 1 and 2 for acceptability [6].

E. Artificial Intelligence Markup Language (AIML)

AIML interpreter tries to match word by word to gain longest pattern matching and try to find which the best one is. This behavior can be described with Graph master set of files and directories containing a set of nodes which is called node master and branches represents first words of all patterns and wildcard symbols [11].

IV. CONCLUSION

Thus, this paper presents survey on intelligent chatbot for guided navigation of a repository content.

ACKNOWLEDGEMENT

This survey is supported by the Sinhgad Academy of Engineering, under the Dept. of Computer Engineering and is supervised by Prof. S.N. Shelke. This Project is sponsored by Prashant Kaushal, (Technical Architect) Konsoftech Solutions Pvt. Ltd.

REFERENCES

- [1] Lili, Z., & Yanli, W. (2010). Design of Harmonious Human-Computer Interaction Based on Intelligence Technology Development. 2010 International Conference on Intelligent System Design and Engineering Application.
- [2] D'Haro, L. F., and amp; Banchs, R. E. (2017). Learning to predict the adequacy of answers in chat-oriented human agent dialogs. TENCON 2017 - 2017 IEEE Region 10 Conference.
- [3] Joseph Weizenbaum. ELIZA a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1):36 45.
- [4] Richard S. Wallace. Be Your Own Botmaster: The Step By Step Guide to Creating, Hosting and Selling Your Own AI ChatBot on Pandorabots. ALICE AI foundations, Incorporated
- [5] Ranoliya, B. R., Raghuvanshi, N., & Singh, S. (2017). Chatbot for university related FAQs. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
- [6] Ly Pichponreay, Jin-Hyuk Kim, Chi-Hwan Choi, Kyung-Hee Lee, & Wan-Sup Cho. (2016). Smart answering Chatbot based on OCR and Overgenerating Transformations and Ranking. 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN).
- [7] Thomas, N. T. (2016). An e-business chatbot using AIML and LSA. 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
- [8] Bhawiyuga, A., Fauzi, M. A., Pramukantoro, E. S., & Yahya, W. (2017). Design of E-commerce chat robot for automatically answering customer question. 2017 International Conference on Sustainable Information Engineering and Technology (SIET).
- [9] Niranjana, M., Saipreethy, M. S., & Kumar, T. G. (2012). An intelligent question answering conversational agent using Naïve Bayesian classifier. 2012 IEEE International Conference on Technology Enhanced Education (ICTEE).
- [10] Shah, A., Jain, B., Agrawal, B., Jain, S., & Shim, S. (2018). Problem solving chatbot for data structures. 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC).
- [11] Satu, M. S., Parvez, M. H., & Shamim-Al-Mamun. (2015). Review of integrated applications with AIML based chatbot. 2015 International Conference on Computer and Information Engineering (ICCIE).